

Optimal Low-Rank Tensor Recovery from Separable Measurements: Four Contractions Suffice

May 18, 2015

Parikshit Shah¹ Nikhil Rao² Gongguo Tang³

Abstract

Tensors play a central role in many modern machine learning and signal processing applications. In such applications, the target tensor is usually of low rank, i.e., can be expressed as a sum of a small number of rank one tensors. This motivates us to consider the problem of low rank tensor recovery from a class of linear measurements called *separable measurements*. As specific examples, we focus on two distinct types of separable measurement mechanisms (a) Random projections, where each measurement corresponds to an inner product of the tensor with a suitable random tensor, and (b) the completion problem where measurements constitute revelation of a random set of entries. We present a computationally efficient algorithm, with rigorous and *order-optimal* sample complexity results (upto logarithmic factors) for tensor recovery. Our method is based on reduction to matrix completion sub-problems and adaptation of Leurgans' method for tensor decomposition. We extend the methodology and sample complexity results to higher order tensors, and experimentally validate our theoretical results

1 Introduction

Tensors provide compact representations for multi-dimensional, multi-perspective data in many problem domains, including image and video processing [50, 33, 25], collaborative filtering [26, 16], statistical modeling [3, 2], array signal processing [32, 41], psychometrics [48, 42], neuroscience [6, 34], and large-scale data analysis [36, 43, 44, 1, 18]. In this paper we consider the problem of tensor recovery - given partial information of a tensor via linear measurements, one wishes to learn the entire tensor. While this inverse problem is ill-posed in general, we will focus on the setting where the underlying tensor is simple. The notion of simplicity that we adopt is based on the (*Kruskal*) *rank* of the tensor, which much like the matrix rank is of fundamental importance - tensors of lower rank have fewer constituent components and are hence simple. For example, video sequences are naturally modeled as tensors, and these third order tensors have low rank as a result of homogeneous variations in the scene [47]. Unlike the matrix case, however, computational tasks related to the tensor rank such as spectral decompositions, rank computation, and regularization are fraught with computational intractability [21, 28] in the worst case.

¹Yahoo! Labs parikshit@yahoo-inc.com

²UT Austin nikhilr@cs.utexas.edu

³Colorado School of Mines gtang@mines.edu

We focus on linear inverse problems involving tensors. Linear measurements of an unknown tensor \mathbf{X} are specified by $y = \mathcal{L}(\mathbf{X})$ where \mathcal{L} is a linear operator and $y \in \mathbb{R}^m$. Here the quantity m refers to the number of measurements, and the minimum number of measurements⁴ m required to reliably recover \mathbf{X} (called the *sample complexity*) is of interest. While in general, such problems are ill-posed and unsolvable when m is smaller than the dimensionality of \mathbf{X} , the situation is more interesting when the underlying signal (tensor) is structured, and the sensing mechanism $\mathcal{L}(\cdot)$ is able to exploit this structure. For instance, similar ill-posed problems are solvable, even if m is substantially lower than the ambient dimension, when the underlying signal is a sparse vector, or a low-rank matrix, provided that $\mathcal{L}(\cdot)$ has appropriate properties.

We focus for the most part on tensors of order 3, and show later that all our results extend to the higher order case in a straightforward way. We introduce a class of measurement operators known as *separable measurements*, and present an algorithm for low-rank tensor recovery for the same. We focus on two specific measurement mechanisms that are special cases of separable mechanisms:

- *Separable random projections:* For tensors of order 3, we consider observations where the i^{th} measurement is of the form $\mathcal{L}_i(\mathbf{X}) := \langle a \otimes A_i, \mathbf{X} \rangle$, where a is a random unit vector, A_i is a random matrix, and \otimes represents an outer product of the two. For higher order tensors, the measurements are defined in an analogous manner. Here $\langle \cdot, \cdot \rangle$ is the tensor inner product (to be made clear in the sequel).
- *Completion:* The measurements here are simply a subset of the entries of the true tensor. The entries need to be restricted to merely four slices of the tensor, and can be random within these slices.

For both the random projection and completion settings, we analyze the performance of our algorithm and prove sample complexity bounds.

The random sampling mechanisms mentioned above are of relevance in practical applications. For instance, the Gaussian random projection mechanism described above is a natural candidate for compressively sampling video and multi-dimensional imaging data. For applications where such data is “simple” (in the sense of low rank), the Gaussian sensing mechanism may be a natural means of compressive encoding.

The completion framework is especially relevant to machine learning applications. For instance, it is useful in the context of multi-task learning [40], where each individual of a collection of inter-related tasks corresponds to matrix completion. Consider the tasks of predicting ratings assigned by users for different clothing items, this is naturally modeled as a matrix completion problem [12]. Similarly, the task of predicting ratings assigned by the same set of users to accessories is another matrix completion problem. The multi-task of jointly predicting the ratings assigned by the users to baskets of items consisting of both clothing items and accessories is a tensor completion problem.

Another application of tensor completion is that of extending the matrix completion framework for contextual recommendation systems. In such a setup, one is given a *rating matrix* that is indexed by users and items, and the entries correspond to the ratings given by different users to different items. Each user provides ratings for only a fraction of the items (these constitute the sensing operator $\mathcal{L}(\cdot)$), and one wishes to infer the ratings for all the others. Assuming that such a rating matrix is low rank is equivalent to assuming the presence of a small number of latent variables that drive the rating process. An interesting twist to this setup which requires a tensor based approach

⁴We use the terms measurements and samples interchangeably.

is *contextual* recommendation - i.e. where different users provide ratings in different contexts (e.g., location, time, activity). Such a setting is naturally modeled via tensors; the three modes of the tensor are indexed by users, items, and contexts. The underlying tensor may be assumed to be low rank to model the small number of latent variables that influence the rating process. In this setting, our approach would need a few samples about users making decisions in two different contexts (this corresponds to two slices of the tensor along the third mode), and enough information about two different users providing ratings in a variety of different contexts (these are two slices along the first mode). Once the completion problem restricted to these slices is solved, one can complete the entire tensor by performing simple linear algebraic manipulations.

Of particular note concerning our algorithm and the performance guarantees are the following:

- **Sample complexity:** In the absence of noise, our algorithm, named T-ReCs (Tensor Recovery via Contractions), provably and exactly recovers the true tensor and achieves an *order-optimal* sample complexity for exact recovery of the underlying tensor in the context of random sensing, and order optimal modulo logarithmic factors in the context of tensor completion. Specifically, for a third order tensor of rank r and largest dimension n , the achieved sample complexity is $O(nr)$ for recovery from separable random projections, and $O(nr \log^2 n)$ for tensor completion. (These correspond to Theorems 3.3 and 3.6 respectively.) More generally, for order K tensors the corresponding sample complexities are $O(Knr)$ and $O(Knr \log^2 n)$ respectively (Theorems 4.4 and 4.5).
- **Factorization:** Equally important is the fact that our method recovers a minimal rank factorization in addition to the unknown tensor. This is of importance in applications such as dimension reduction and also latent variable models [2] involving tensors where the factorization itself holds meaningful interpretational value.
- **Absence of strong assumptions:** Unlike some prior art, our analysis relies only on relatively weak assumptions - namely that the rank of the tensor be smaller than the (smallest) dimension, that the factors in the rank decomposition be linearly independent, non-degenerate, and (for the case of completion) other standard assumptions such as incoherence between the factors and the sampling operator. We do not, for instance, require orthogonality-type assumptions of the said factors, as is the case in [2, 23].
- **Computational efficiency:** Computationally, our algorithm essentially reduces to linear algebraic operations and the solution of matrix nuclear norm (convex) optimization subproblems, and is hence extremely tractable. Furthermore, our nuclear norm minimization methods deal with matrices that are potentially much smaller, up to factors of n , than competing methods that “matricize” the tensor via unfolding [35, 46]. In addition to recovering the true underlying tensor, it also produces its unique rank decomposition.
- **Simplicity:** Our algorithm is conceptually simple - both to implement as well as to analyze. Indeed the algorithm and its analysis follow in a transparent manner from Leurgans’ algorithm (a simple linear algebraic approach for tensor decomposition) and standard results for low-rank matrix recovery and completion. We find this intriguing, especially considering the “hardness” of most tensor problems [21, 28]. Recent work in the area of tensor learning has focused on novel regularization schemes and algorithms for learning low rank tensors; the proposed approach potentially obviates the need for developing these in the context of separable measurements.

The fundamental insight in this work is that while solving the tensor recovery problem directly may seem challenging (for example we do not know of natural tractable extensions of the “nuclear norm” for tensors), very important information is encoded in a two-dimensional matrix “sketch” of the tensor which we call a contraction. (This idea seems to first appear in [31], and is expanded upon in [7, 20] in the context of tensor decomposition.) These sketches are formed by taking linear combinations of two-dimensional slices of the underlying tensor - indeed the slices themselves may be viewed as “extremal” contractions. For the Gaussian random projections case, the contractions will be random linear combinations of slices, whereas for the completion setting the contractions we work with will be the slices themselves, randomly subsampled. Our method focuses on recovering these contractions efficiently (using matrix nuclear norm regularization) as a first step, followed by additional processing to recover the true tensor.

1.1 Related Work and Key Differences

With a view to computational tractability, the notion of *Tucker rank* of a tensor has been explored; this involves *matricizations* along different modes of the tensor and the ranks of the associated matrices. Based on the idea of Tucker rank, Tomioka et al. [46] have proposed and analyzed a nuclear norm heuristic for tensor completion, thereby bringing tools from matrix completion [12] to bear for the tensor case. Mu et al. [35], have extended this idea further by studying reshaped versions of tensor matricizations. However, to date, the sample complexity associated to matrix-based regularization seem to be orders far from the anticipated sample complexity (for example based on a count of the degrees of freedom in the problem) [35]. In this paper we resolve this conundrum by providing an efficient algorithm that provably enjoys order optimal sample complexity in the order, dimension, and rank of the tensor.

In contrast to the matricization approach, alternative approaches for tensor completion with provable guarantees have appeared in the literature. In the restricted setting when the tensor has a symmetric factorization [5] (in contrast we are able to work in the general non-symmetric setting), the authors propose employing the Lasserre hierarchy via a semidefinite programming based approach. Unfortunately, the method proposed in [5] is not scalable - it requires solving optimization problems at the 6th level of the Lasserre hierarchy which makes solving even moderate-sized problems numerically impractical as the resulting semidefinite programs grow rapidly with the dimension. Furthermore, the guarantees provided in [5] are of a different flavor - they provide error bounds in the noisy setting, whereas we provide exact recovery results in the noiseless setting. Alternate methods based on thresholding in the noisy setting have also been studied in [4]. An alternating minimization approach for tensor completion was proposed in [23]. Their approach relies on the restrictive assumptions also - that the underlying tensor be symmetric and orthogonally decomposable (we make no such assumptions), and neither do the sample complexity bounds scale optimally with the dimensions or the rank. Unlike alternating minimization schemes that are efficient but rely on careful initializations, our method directly solves convex optimization programs followed by linear algebraic manipulations. Also relevant is [49], where the authors propose solving tensor completion using the tensor nuclear norm regularizer; this approach is not known to be computationally tractable (no polynomial time algorithm is known for minimizing the tensor nuclear norm) and the guarantees they obtain do not scale optimally with the dimension and rank. Finally a method based on the tubal rank and *t*-SVD of a tensor [51] has also recently been proposed, however the sample complexity does not scale optimally. As a final point of contrast to the aforementioned work, our method is also conceptually very simple - both to implement and analyze.

In Table 1.1, we provide a brief comparison of the relevant approaches, their sample complexities in both the third order and higher order settings as well as a few key features of each approach.

Reference	Sample Complexity (3^{rd} order)	Sample Complexity (K^{th} order)	Key Features
[46]	$O(rn^2)$	$O(rn^{K-1})$	Tucker rank, tensor unfolding
[35]	$O(rn^2)$	$O(rn^{\lfloor \frac{K}{2} \rfloor})$	Tucker rank, tensor unfolding
[23]	$O(r^5 n^{\frac{3}{2}} \log^5 n)$	-	Kruskal rank, alternating minimization, orthogonally decomposable tensors, symmetric setting, completion only.
[51]	$O(rn^2 \log n)$	-	Tensor tubal rank, completion only
[49]	$O(r^{\frac{1}{2}} (n \log n)^{\frac{3}{2}})$	$O(n^{\frac{K}{2}} \text{polylog}(n))$	Kruskal rank, Exact tensor nuclear norm minimization, computationally intractable, completion only.
Our Method	$O(nr)$ (random projection) $O(nr \log^2 n)$ (completion)	$O(Knr)$ (random projection) $O(Knr \log^2 n)$ (completion)	Kruskal rank, separable measurements, Leurgans' algorithm

Table 1: Table comparing sample complexities of various approaches.

The rest of the paper is organized as follows: in Section 2, we introduce the problem setup and describe the approach and result in the most general setting. We also describe Leurgans' algorithm, an efficient linear algebraic algorithm for tensor decomposition, which our results build upon. In Section 3 we specialize our results for both the random projections and the tensor completion cases. We extend these results and our algorithm to higher order tensors in Section 4. We perform experiments that validate our theoretical results in Section 5. In Section 6, we conclude the paper and outline future directions.

2 Approach and Basic Results

In this paper, vectors are denoted using lower case characters (e.g. x, y, a, b , etc.), matrices by upper-case characters (e.g. X, Y , etc.) and tensors by upper-case bold characters (e.g. $\mathbf{X}, \mathbf{T}, \mathbf{A}$ etc.). Given two third order tensors \mathbf{A}, \mathbf{B} , their inner product is defined as:

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j,k} A_{ijk} B_{ijk}.$$

The Euclidean norm of a tensor \mathbf{A} is generated by this inner product, and is a straightforward extension of the matrix Frobenius norm:

$$\|\mathbf{A}\|_F^2 := \langle \mathbf{A}, \mathbf{A} \rangle.$$

We will work with tensors of third order (representationally to be thought of as three-way arrays), and the term mode refers to one of the axes of the tensor. A slice of a tensor refers to a two dimensional matrix generated from the tensor by varying indices along two modes while keeping the third mode fixed. For a tensor \mathbf{X} we will refer to the indices of the i^{th} mode-1 slice (i.e., the slice corresponding to the indices $\{i\} \times [n_2] \times [n_3]$) by $S_i^{(1)}$, where $[n_2] = \{1, 2, \dots, n_2\}$ and $[n_3]$ is defined similarly. We denote the matrix corresponding to $S_i^{(1)}$ by X_i^1 . Similarly the indices of the k^{th} mode-3 slice will be denoted by $S_k^{(3)}$ and the matrix by X_k^3 .

Given a tensor of interest \mathbf{X} , consider its decomposition into rank one tensors

$$\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i, \quad (1)$$

where $\{u_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_1}$, $\{v_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_2}$, and $\{w_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_3}$. Here \otimes denotes the tensor product, so that $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a tensor of order 3 and dimension $n_1 \times n_2 \times n_3$. Without loss of generality, throughout this paper we assume that $n_1 \leq n_2 \leq n_3$. We will first present our results for third order tensors, and analogous results for higher orders follow in a transparent manner. We will be dealing with *low-rank* tensors, i.e. those tensors with $r \leq n_1$. Tensors can have rank larger than the dimension, indeed $r \geq n_3$ is an interesting regime, but far more challenging and will not be dealt with here.

Kruskal's Theorem [29] guarantees that tensors satisfying Assumption 2.1 below have a unique minimal decomposition into rank one terms of the form (1). The minimal number of terms is called the (Kruskal) rank⁵ of the tensor \mathbf{X} .

Assumption 2.1. *The sets $\{u_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_1}$ and $\{v_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_2}$ are sets of linearly independent vectors and the set $\{w_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_3}$ is a set of pairwise independent vectors*

While rank decomposition of tensors in the worst case is known to be computationally intractable [21], it is known that the (mild) assumption stated in Assumption 2.1 above suffices for an algorithm known as Leurgans' algorithm [31, 7] to correctly identify the factors in this unique decomposition. In this paper, we will work with the following, somewhat stronger assumption:

Assumption 2.2. *The sets $\{u_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_1}$, $\{v_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_2}$, and $\{w_i\}_{i=1,\dots,r} \subseteq \mathbb{R}^{n_3}$ are sets of linearly independent vectors.*

⁵The Kruskal rank is also known as the CP rank in the literature.

2.1 Separable Measurement Mechanisms

As indicated above in the preceding discussions, we are interested in tensor linear inverse problems where, given measurements of the form $y_i = \mathcal{L}_i(\mathbf{X})$ $i = 1, 2, \dots, m$, we recover the unknown tensor \mathbf{X} . We focus on a class of measurement mechanisms $\mathcal{L}(\cdot)$ which have a special property which we call *separability*. We define the notion of separable measurements formally:

Definition 2.1. Consider a linear operator $\mathcal{L} : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^n$. We say that \mathcal{L} is separable with respect to the third mode if there exist $w \in \mathbb{R}^{n_3}$ and a linear operator $\mathcal{T} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^n$, such that for every $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$:

$$\mathcal{L}(\mathbf{X}) = \sum_{i=1}^{n_3} w_i \mathcal{T}(X_i^3).$$

This definition extends in a natural way for separability of operators with respect to the second and first modes. In words, separability means that the effect of the linear operator $\mathcal{L}(\cdot)$ on a tensor can be decomposed into the (weighted) sum of actions of a *single* linear operator $\mathcal{T}(\cdot)$ acting on *slices* of the tensor along a particular mode.

In several applications involving inverse problems, the design of appropriate measurement mechanisms is itself of interest. Indeed sensing methods that lend themselves to recovery from a small number of samples via efficient computational techniques has been intensely studied in the signal processing, compressed sensing, and machine learning literature [13, 10, 39, 38, 19, 45]. In the context of tensors, we argue, separability of the measurement operator is a desirable property for precisely these reasons; because it lends itself to recovery up to almost optimal sample complexity via scalable computational methods (See for example [30] for rank one measurement operators in the matrix case). We now describe a few interesting measurement mechanisms that are separable.

1. **Separable random projections:** Given a matrix $M \in \mathbb{R}^{n_1 \times n_2}$ and vector $v \in \mathbb{R}^{n_3}$, we define the following two notions of “outer products” of M and v :

$$[M \otimes v]_{ijk} := M_{ij}v_k \quad [v \otimes M]_{ijk} := v_i M_{jk}.$$

Hence, the k^{th} mode 3 slice of the tensor $M \otimes v$ is the matrix $v_k M$. Similarly, the i^{th} mode 1 slice of the tensor $v \otimes M$ is the matrix $v_i M$.

A typical random separable projection is of the form:

$$\mathcal{L}(\mathbf{X}) = \begin{bmatrix} \langle A_1 \otimes a, \mathbf{X} \rangle \\ \vdots \\ \langle A_m \otimes a, \mathbf{X} \rangle \end{bmatrix} \quad (2)$$

where $A_i \in \mathbb{R}^{n_1 \times n_2}$ is a random matrix drawn from a suitable ensemble such as the Gaussian ensemble with each entry drawn independently and identically from $\mathcal{N}(0, 1)$, and $a \in \mathbb{R}^{n_3}$ is also a random vector, for instance distributed uniformly on the unit sphere in n_3 dimensions (i.e. with each entry drawn independently and identically from $\mathcal{N}(0, 1)$ and then suitably normalized).

To see that such measurements are separable, note that:

$$\langle A_i \otimes a, \mathbf{X} \rangle = \sum_{k=1}^{n_3} a_k \langle A_i, X_k^3 \rangle,$$

so that the operator $\mathcal{T}(\cdot)$ from Definition 2.1 in this case is simply given by:

$$\mathcal{T}(X) = \begin{bmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{bmatrix}.$$

Random projections are of basic interest in signal processing, and have played a key role in the development of sparse recovery and low rank matrix recovery literature [39, 10]. From an application perspective they are relevant because they provide a method of compressive and lossless coding of “simple signals” such as sparse vectors [10] and low rank matrices [39]. In subsequent sections we will establish that separable random projections share this desirable feature for low-rank tensors.

2. **Tensor completion:** In tensor completion, a subset of the entries of the tensor \mathbf{X} are revealed. Specifically, given a tensor \mathbf{X} , a subset of the entries \mathbf{X}_{ijk} for $i, j, k \in \Omega$ are revealed for some index set $\Omega \subseteq [n_1] \times [n_2] \times [n_3]$ (we denote this by $(\mathbf{X})_\Omega$). Whether or not the measurements are separable depends upon the nature of the set Ω . For the i^{th} mode-1 slice let us define

$$\Omega_i^{(1)} := \Omega \cap S_i^{(1)} \quad m_i^{(1)} := |\Omega \cap S_i^{(1)}|.$$

Measurements derived from entries within a *single slice* of the tensor are separable. This follows from the fact that for $\mathcal{L}(\mathbf{X}) := (\mathbf{X})_{\Omega_i^{(1)}}$, we have:

$$\mathcal{L}(\mathbf{X}) = \sum_{j=1}^{n_1} (\delta_i)_j \mathcal{M}_{\Omega_i^{(1)}}(X_j^{(1)})$$

where $\delta_i \in \mathbb{R}^{n_1}$ is a vector with a one in the i index and zero otherwise, and \mathcal{M}_Ω is the operator that acts on a matrix X , extracts the indices corresponding to the index Ω , and returns the resulting vector. Comparing to Definition 2.1, we have $w = \delta_i$ and $\mathcal{T} = \mathcal{M}_{\Omega_i^{(1)}}$. As a trivial extension, measurements obtained from parallel slices where the index set restricted to these slices is identical are also separable.

Analogous to matrix completion, tensor completion is an important problem due to its applications to machine learning; the problems of multi-task learning and contextual recommendation are both naturally modeled in this framework as described in Section 1.

3. **Rank one projections** Another separable sensing mechanism of interest is via rank-one projections of the tensor of interest. Specifically, measurements of the form:

$$\mathcal{L}(\mathbf{X}) = \begin{bmatrix} \langle a_1 \otimes b_1 \otimes c, \mathbf{X} \rangle \\ \vdots \\ \langle a_m \otimes b_m \otimes c, \mathbf{X} \rangle \end{bmatrix}$$

are also separable. Mechanisms of this form have recently gained interest in the context of low rank (indeed rank-one) matrices due to their appearance in the context of phase retrieval problems [14] and statistical estimation [30, 9]. We anticipate that studying rank one projections in the context of tensors will give rise to interesting applications in a similar spirit.

4. **Separable sketching** The notion of covariance sketching (and more generally, matrix sketching) [19] allows for the possibility of compressively acquiring a matrix X via measurements $Y = AXB^T$, where $A \in \mathbb{R}^{m_1 \times p}$ and $B \in \mathbb{R}^{m_2 \times p}$, $X \in \mathbb{R}^p$, and $m_1, m_2 < p$. The problem of recovering X from such measurements is of interest in various settings such as when one is interested in recovering a covariance matrix from compressed sample paths, and graph compression [19]. In a similar spirit, we introduce the notion of separable sketching of tensors defined via:

$$Y_{qs} = \mathcal{L}(\mathbf{X}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} A_{qi} B_{sj} c_k \mathbf{X}_{ijk}.$$

In the above $A \in \mathbb{R}^{m_1 \times n_1}$, $B \in \mathbb{R}^{m_2 \times n_2}$, $c \in \mathbb{R}^{n_3}$, and $Y \in \mathbb{R}^{m_1 \times m_2}$. Note that $\mathcal{T}(Z) = AZB^T$, i.e. precisely a matrix sketch of tensor slices. The problem of recovering \mathbf{X} from Y is thus a natural extension of matrix sketching to tensors.

Finally, we note that while a variety of separable sensing mechanisms are proposed above, many sensing mechanisms of interest are not separable. For instance, a measurement of the form $\mathcal{L}(\mathbf{X}) = \langle \mathbf{A}, \mathbf{X} \rangle$ where \mathbf{A} is a full rank tensor is *not* separable. Similarly, completion problems where entries of the tensor are revealed randomly and uniformly throughout the tensor (as apposed to from a single slice) are also not separable (although they may be thought of as a union of separable measurements). In Section 3, we will provide sample complexity bounds for exact recovery for the first two aforementioned measurement mechanisms (i.e. random projections and tensor completion); the arguments extend in a natural manner to other separable sensing mechanisms.

2.1.1 Diversity in the Measurement Set

In order to recover the low rank tensor from a few measurements using our algorithm, we need the set of measurements to be a *union* of separable measurements which satisfy the following:

1. **Diversity across modes:** Measurements of the form (8) are separable with respect to the third mode. For the third order case, we also need an additional set of measurements separable with respect to the first mode ⁶. This extends naturally also to the higher order case.
2. **Diversity across separable weights:** Recalling the notion of weight vectors, w , from Definition: 2.1, we require that for both modes 1 and 3, each mode has two distinct sets of separable measurements with distinct weight vectors.

To make the second point more precise later, we introduce the formal notation we will use in the rest of the paper for the measurement operators:

$$y_k^{(i)} = \mathcal{L}_k^{(i)}(\mathbf{X}) = \sum_{j=1}^{n_3} \left(w_k^{(i)} \right)_j \mathcal{T}_k^{(i)}(X_j^3)$$

In the above, the index $i \in \{1, 3\}$ refers to the mode with respect to which that measurement is separable. For each mode, we have two distinct sets of measurements corresponding to two different weight vectors $w_k^{(i)}$, with $k \in \{1, 2\}$. For each k and i , we may have potentially different operators

⁶Any two modes suffice. In this paper we will focus on separability w.r.t the first and third modes.

$\mathcal{T}_k^{(i)}$ (though they need not be different). To simplify notation, we will subsequently assume that $\mathcal{T}_1^{(i)} = \mathcal{T}_2^{(i)} = \mathcal{T}^{(i)}$. Collectively, all these measurements will be denoted by:

$$y = \mathcal{L}(\mathbf{X}),$$

where it is understood that y is a concatenation of the vectors $y_k^{(i)}$ and similarly $\mathcal{L}(\cdot)$ is a concatenation of $\mathcal{L}_k^{(i)}(\cdot)$. We will see in the subsequent sections that when we have diverse measurements across different modes and different weight vectors, and when the $\mathcal{T}^{(i)}$ are chosen suitably, one can efficiently recover an unknown tensor from an (almost) optimal number of measurements of the form $y = \mathcal{L}(\mathbf{X})$.

2.2 Tensor Contractions

A basic ingredient in our approach is the notion of a tensor contraction. This notion will allow us to form a bridge between inverse problems involving tensors and inverse problems involving matrices, thereby allowing us to use matrix-based techniques to solve tensor inverse problems.

For a tensor \mathbf{X} , we define its mode-3 *contraction* with respect to a contraction vector $a \in \mathbb{R}^{n_3}$, denoted by $X_a^3 \in \mathbb{R}^{n_1 \times n_2}$, as the following matrix:

$$[X_a^3]_{ij} = \sum_{k=1}^{n_3} \mathbf{X}_{ijk} a_k, \quad (3)$$

so that the resulting matrix is a weighted sum of the mode-3 slices of the tensor \mathbf{X} . We similarly define the mode-1 contraction with respect to a vector $c \in \mathbb{R}^{n_1}$ as

$$[X_c^1]_{jk} = \sum_{i=1}^{n_1} \mathbf{X}_{ijk} c_i, \quad (4)$$

Note that when $a = e_k$, a standard unit vector, $X_a^3 = X_k^3$, i.e. a tensor slice. We will primarily be interested in two notions of contraction in this paper:

- *Random Contractions*, where a is a random vector distributed uniformly on the unit sphere. These will play a role in our approach for recovery from random projections.
- *Coordinate Contractions*, where a is a canonical basis vector, so that the resulting contractions is a tensor slice. These will play a role in our tensor completion approach.

We now state a basic result concerning tensor contractions.

Lemma 2.1. *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, with $n_1 \leq n_2 \leq n_3$ be a tensor of rank $r \leq n_1$. Then the rank of X_a^3 is at most r . Similarly, if $r \leq \min\{n_2, n_3\}$ then the rank of X_c^1 is at most r .*

Proof. Consider a tensor $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$. The reader may verify in a straightforward manner that X_a^3 enjoys the decomposition:

$$X_a^3 = \sum_{i=1}^r \langle w_i, a \rangle u_i v_i^T. \quad (5)$$

The proof for the rank of X_c^1 is analogous. □

Note that while (5) is a matrix decomposition of the contraction, it is not a singular value decomposition (the components need not be orthogonal, for instance). Indeed it does not seem “canonical” in any sense. Hence, given contractions, resolving the components is a non-trivial task.

A particular form of degeneracy we will need to avoid is situations where $\langle w_i, a \rangle = 0$ for (5). It is interesting to examine this in the context of coordinate contractions, i.e. when $a = e_k$, we have $X_{e_k}^3 = X_3^k$ (i.e. the k^{th} mode 3 slice), by Lemma 2.1, we see that the tensor slices are also of rank at most r . Applying $a = e_k$ in the decomposition (5), we see that if for some vector $w_i \in \mathbb{R}^{n_3}$ in the above decomposition we have that the k^{th} component of w_i (i.e. $(w_i)_k$) is zero then $\langle w_i, e_k \rangle = 0$, and hence this component is missing in the decomposition of X_k^3 . As a consequence the rank of X_k^3 drops, and in a sense information about the factors u_k, v_k is “lost” from the contraction. We will want to avoid such situations and thus introduce the following definition:

Definition 2.2. Let $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$. We say that the contraction X_a^3 is non-degenerate if $\langle w_i, a \rangle \neq 0$, for all $i = 1, \dots, r$.

We will extend the terminology and say that the tensor \mathbf{X} is *non-degenerate* at mode 3 and component k if the k^{th} tensor slice is non-degenerate, i.e. component k of the vectors $w_i, i = 1, \dots, r$ are all non-zero. The above definition extends in a natural way to other modes and components. The non-degeneracy condition is trivially satisfied (almost surely) when:

1. The vector a with respect to which the contraction is computed is suitably random, for instance random normal. In such situations, non-degeneracy holds almost surely.
2. When $a = e_k$ (i.e. the contraction is a slice), and the tensor factors are chosen from suitable random ensembles, e.g. when the low rank tensors are picked such that the rank one components u_i, v_i, w_i are Gaussian random vectors, or random orthogonal vectors ⁷.

We will also need the following definition concerning the genericity of a pair of contractions:

Definition 2.3. Given a tensor $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$, a pair of contractions X_a^3, X_b^3 are pairwise generic if the diagonal entries of the (diagonal) $D_a D_b^{-1}$ are all distinct, where $D_a = \text{diag}(\langle w_1, a \rangle, \dots, \langle w_r, a \rangle)$, $D_b = \text{diag}(\langle w_1, b \rangle, \dots, \langle w_r, b \rangle)$.

Remark We list two cases where pairwise genericity conditions hold in this paper.

1. In the context of random contractions, for instance when the contraction vectors a, b are sampled uniformly and independently on the unit sphere. In this case pairwise genericity holds almost surely.
2. In the context of tensor completion where $a = e_{k_1}, b = e_{k_2}$, the two diagonal matrices $D_a = \text{diag}((w_1)_{k_1}, \dots, (w_r)_{k_1})$, and $D_b = \text{diag}((w_1)_{k_2}, \dots, (w_r)_{k_2})$. Thus the pairwise genericity condition is a genericity requirement of the tensor factors themselves, namely that the ratios $\frac{(w_i)_{k_1}}{(w_i)_{k_2}}$ all be distinct for $i = 1, \dots, r$. We will abuse terminology, and call such a tensor pairwise generic with respect to mode 3 slices k_1, k_2 . This form of genericity is easily seen to hold, for instance when the tensor factors are drawn from suitable random ensembles such as random normal and random uniformly distributed on the unit sphere.

⁷the latter is known as random orthogonal model in the matrix completion literature [38]

The next lemma, a variation of which appears in [7, 31] shows that when the underlying tensor is non-degenerate, it is possible to decompose a tensor from pairwise generic contractions.

Lemma 2.2. [7, 31] *Suppose we are given an order 3 tensor $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$ of size $n_1 \times n_2 \times n_3$ satisfying the conditions of Assumption 2.1. Suppose the contractions X_a^3 and X_b^3 are non-degenerate, and consider the matrices M_1 and M_2 formed as:*

$$M_1 = X_a^3 (X_b^3)^\dagger \quad M_2 = (X_b^3)^\dagger X_a^3.$$

Then the eigenvectors of M_1 (corresponding to the non-zero eigenvalues) are $\{u_i\}_{i=1,\dots,r}$, and the eigenvectors of M_2^T are $\{v_i\}_{i=1,\dots,r}$.

Proof. Suppose we are given an order 3 tensor $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. From the definition of contraction (3), it is straightforward to see that

$$X_a^3 = U D_a V^T \quad D_a = \text{diag}(a^T w_1, \dots, a^T w_r)$$

$$X_b^3 = U D_b V^T \quad D_b = \text{diag}(b^T w_1, \dots, b^T w_r).$$

In the above decompositions, $U \in \mathbb{R}^{n_1 \times r}$, $V \in \mathbb{R}^{n_2 \times r}$, and the matrices $D_a, D_b \in \mathbb{R}^{r \times r}$ are diagonal and non-singular (since the contractions are non-degenerate). Now,

$$\begin{aligned} M_1 &:= X_a^3 (X_b^3)^\dagger \\ &= U D_a V^T (V^\dagger)^T D_b^{-1} U^\dagger \\ &= U D_a D_b^{-1} U^\dagger \end{aligned} \tag{6}$$

and similarly we obtain

$$M_2^T = V D_b^{-1} D_a V^\dagger. \tag{7}$$

Since we have $M_1 U = U D_a D_b^{-1}$ and $M_2^T V = V D_b^{-1} D_a$, it follows that the columns of U and V are eigenvectors of M_1 and M_2^T respectively (with corresponding eigenvalues given by the diagonal matrices $D_a D_b^{-1}$ and $D_b^{-1} D_a$). \square

Remark Note that while the eigenvectors $\{u_i\}, \{v_j\}$ are thus determined, a source of ambiguity remains. For a fixed ordering of the u_i one needs to determine the order in which the v_j are to be arranged. This can be (generically) achieved by using the (common) eigenvalues of M_1 and M_2 for pairing. If the contractions X_a^3, X_b^3 satisfy pairwise genericity, we see that the diagonal entries of the matrix $D_a D_b^{-1}$ are distinct. It then follows that the eigenvalues of M_1, M_2 are distinct, and can be used to pair the columns of U and V .

2.3 Leurgans' algorithm

We now describe Leurgans' algorithm for tensor decomposition in Algorithm 1. In the next section, we build on this algorithm to solve tensor inverse problems to obtain optimal sample complexity bounds. In words, Algorithm 1 essentially turns a problem involving decomposition of tensors into that of decomposition of matrices. This is achieved by first computing mode 3 contractions of the given tensor \mathbf{X} with respect to two non-degenerate and pairwise generic vectors a, b (e.g. randomly uniformly distributed on the unit sphere). Given these contractions, one can compute matrices M_1 and M_2 as described in Lemma 2.2 whose eigenvectors turn out to be *precisely* (up to scaling) the

vectors u_i and v_i of the required decomposition. Finally the w_i can be obtained by inverting an (overdetermined) system of linear equations, giving a unique and exact solution.

The correctness of the algorithm follows directly from Lemma 2.2.

In this paper, we extend this idea to solving ill-posed linear inverse problems of tensors. The key idea is that since the contractions preserve information about the tensor factors, we focus on recovering the contractions first. Once those are recovered, we simply need to compute eigendecompositions to recover the factors themselves.

Algorithm 1 Leurgans’ algorithm for tensor decomposition

- 1: **Input:** Tensor \mathbf{X}
 - 2: Generate contraction vectors $a, b \in \mathbb{R}^{n_3}$ (such that non-degeneracy and pairwise genericity holds).
 - 3: Compute mode 3 contractions X_a^3 and X_b^3 respectively.
 - 4: Compute eigen-decomposition of $M_1 := X_a^3(X_b^3)^\dagger$ and $M_2 := (X_b^3)^\dagger X_a^3$. Let U and V denote the matrices whose columns are the eigenvectors of M_1 and M_2^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_1 and M_2 .) The eigenvectors, thus arranged are denoted as $\{u_i\}_{i=1,\dots,r}$ and $\{v_i\}_{i=1,\dots,r}$.
 - 5: Solve for w_i in the (over-determined) linear system $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i, i = 1, \dots, m$.
 - 6: **Output:** Decomposition $\mathbf{X} = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$.
-

Remark Note that in the last step, instead of solving a linear system of equations to obtain the w_i , there is an alternative approach whereby one may compute mode 1 contractions and then obtain the factors v_i and w_i . However, there is one minor caveat. Suppose we denote the factors obtained from the modal contractions X_a^3 and X_b^3 by U and V_1 (we assume that these factors are normalized, i.e. the columns have unit Euclidean norm). Now, we can repeat the procedure with two more random vectors c, d to compute the contractions X_c^1 and X_d^1 . We can perform similar manipulations to construct matrices whose eigenvectors are the tensor factors of interest, and thence obtain (normalized) factors V_2 and W . While V_1 and V_2 essentially correspond to the same factors, the matrices themselves may (i) have their columns in different order, and (ii) have signs reversed relative to each other. Hence, while the modal contractions preserve information about the tensor factors, they may need to be properly aligned by rearranging the columns and performing sign reversals, if necessary.

2.4 High Level Approach

The key observation driving the methodology concerns the separability of the measurements. Given a set of separable measurements $y = \mathcal{L}(\mathbf{X})$, from the definition of separability we have:

$$y = \mathcal{L}(\mathbf{X}) = \sum_{i=1}^{n_3} w_i \mathcal{T}(X_i^3) = \mathcal{T}\left(\sum_{i=1}^{n_3} w_i X_i^3\right) = \mathcal{T}(X_w^3).$$

In words, each separable measurement \mathcal{L} acting on the tensor can also be interpreted as a measurement \mathcal{T} acting on a *contraction* of the tensor. Since these contractions are low rank (Lemma 2.1), when the underlying tensor is low-rank, the following nuclear norm minimization problem represents a principled, tractable heuristic for recovering the contraction:

$$\text{minimize}_Z \|Z\|_* \quad \text{subject to} \quad y = \mathcal{T}(Z).$$

Let us informally define \mathcal{T} to be “faithful” if nuclear norm minimization succeeds in exactly recovering the tensor contractions. Provided we correctly recover two contractions each along modes 1 and 3, and furthermore these contractions are non-degenerate and pairwise generic, we can apply Leurgans’ algorithm to the recovered contractions to exactly recover the unknown tensor. This yields the following meta-theorem:

Meta-Theorem. *Given a low rank tensor \mathbf{X} and separable measurements*

$$y_k^{(i)} = \mathcal{L}_k^{(i)}(\mathbf{X}) = \sum_{j=1}^{n_3} \left(w_k^{(i)} \right)_j \mathcal{T}^{(i)}(X_j^3), \quad i \in \{1, 3\}, k \in \{1, 2\}.$$

Suppose the $\mathcal{T}^{(i)}$ are faithful and for the vectors $w_k^{(i)}$, the contractions $X_{w_k^{(i)}}^i$ are non-degenerate and pairwise generic. Then the proposed approach succeeds in exactly recovering the unknown tensor.

In the next section, we will make the above meta-theorem more precise, and detail the precise sample complexities for the separable random projections and tensor completion settings. We will see that faithfulness, non-degeneracy and pairwise genericity hold naturally in these settings.

3 Sample Complexity Results: Third Order Case

3.1 Tensor Recovery via Contractions

We start by describing the main algorithm of this paper more precisely: Tensor Recovery via Contractions (T-ReCs). We assume that we are given separable measurements $y_1^{(3)} = \mathcal{L}_1^{(3)}(\mathbf{X})$, $y_2^{(3)} = \mathcal{L}_2^{(3)}(\mathbf{X})$, $y_1^{(1)} = \mathcal{L}_1^{(1)}(\mathbf{X})$, $y_2^{(1)} = \mathcal{L}_2^{(1)}(\mathbf{X})$. We further assume that the measurements are separable as:

$$\begin{aligned} \mathcal{L}_1^{(3)}(\mathbf{X}) &= \sum_{i=1}^{n_3} a_i \mathcal{T}^{(3)}(X_i^3) & \mathcal{L}_2^{(3)}(\mathbf{X}) &= \sum_{i=1}^{n_3} b_i \mathcal{T}^{(3)}(X_i^3) \\ \mathcal{L}_1^{(1)}(\mathbf{X}) &= \sum_{i=1}^{n_1} c_i \mathcal{T}^{(1)}(X_i^1) & \mathcal{L}_2^{(1)}(\mathbf{X}) &= \sum_{i=1}^{n_1} d_i \mathcal{T}^{(1)}(X_i^1). \end{aligned} \tag{8}$$

where a, b, c, d and $\mathcal{T}^{(3)}$ and $\mathcal{T}^{(1)}$ are known in advance. Given these measurements our algorithm will involve the solution of the following convex optimization problems.

$$\begin{aligned} \underset{Z_1}{\text{minimize}} \quad & \|Z_1\|_* \quad \text{s.t.} \quad y_1^{(3)} = \mathcal{T}^{(3)}(Z_1) \end{aligned} \tag{9}$$

$$\begin{aligned} \underset{Z_2}{\text{minimize}} \quad & \|Z_2\|_* \quad \text{s.t.} \quad y_2^{(3)} = \mathcal{T}^{(3)}(Z_2) \end{aligned} \tag{10}$$

$$\begin{aligned} \underset{Z_3}{\text{minimize}} \quad & \|Z_3\|_* \quad \text{s.t.} \quad y_1^{(1)} = \mathcal{T}^{(1)}(Z_3) \end{aligned} \tag{11}$$

$$\begin{aligned} \underset{Z_4}{\text{minimize}} \quad & \|Z_4\|_* \quad \text{s.t.} \quad y_2^{(1)} = \mathcal{T}^{(1)}(Z_4) \end{aligned} \tag{12}$$

Efficient computational methods have been extensively studied in recent years for solving problems of this type [24]. These matrices form the “input matrices” in the next step which is an

adaptation of Leurgans' method. In this step we form eigendecompositions to reconstruct first the pair of factors u_i, v_i , and then the pairs v_i, w_i (the factors are normalized). Once these are recovered the last step involves solving a linear system of equations for the weights λ_i in

$$\mathcal{L}_1^{(3)}(\mathbf{X}) = \sum_{i=1}^r \lambda_i \mathcal{L}_1^{(3)}(u_i \otimes v_i \otimes w_i) = y_1^{(3)}$$

The pseudocode for T-ReCs is detailed in Algorithm 2.

Algorithm 2 Tensor-Recovery via Contractions (T-ReCs)

- 1: **Input:** Separable measurements $y_1^{(3)} = \mathcal{L}_1^{(3)}(\mathbf{X})$, $y_2^{(3)} = \mathcal{L}_2^{(3)}(\mathbf{X})$, $y_1^{(1)} = \mathcal{L}_1^{(1)}(\mathbf{X})$, $y_2^{(1)} = \mathcal{L}_2^{(1)}(\mathbf{X})$.
- 2: Solve convex optimization problems (9) and (10) to obtain optimal solutions Z_1^* and Z_2^* respectively.
- 3: Compute eigen-decomposition of $M_1 := Z_1^*(Z_2^*)^\dagger$ and $M_2 := (Z_2^*)^\dagger Z_1$. Let U and V denote the matrices whose columns are the eigenvectors of M_1 and M_2^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_1 and M_2 .) The eigenvectors, thus arranged are denoted as $\{u_i\}_{i=1,\dots,r}$ and $\{v_i\}_{i=1,\dots,r}$.
- 4: Solve convex optimization problems (11) and (12) to obtain optimal solutions Z_3^* and Z_4^* respectively.
- 5: Compute eigen-decomposition of $M_3 := Z_3^*(Z_4^*)^\dagger$ and $M_4 := (Z_4^*)^\dagger Z_3$. Let \tilde{V} and \tilde{W} denote the matrices whose columns are the eigenvectors of M_3 and M_4^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_3 and M_4 .) The eigenvectors, thus arranged are denoted as $\{\tilde{v}_k\}_{k=1,\dots,r}$ and $\{\tilde{w}_k\}_{k=1,\dots,r}$.
- 6: Simultaneously reorder the columns of \tilde{V}, \tilde{W} , also performing simultaneous sign reversals as necessary so that the columns of V and \tilde{V} are equal, call the resulting matrix W with columns $\{w_i\}_{i=1,\dots,r}$.
- 7: Solve for λ_i in the (over-determined) linear system

$$y_i = \sum_{i=1}^r \lambda_i \mathcal{L}(u_i \otimes v_i \otimes w_i).$$

- 8: **Output:** Recovered tensor $\mathbf{X} = \sum_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i$.
-

We now focus on the case of recovery from random Gaussian measurements, and then move on to the case of recovery from partially observed samples - in these situations not only are the measurements separable but one can also obtain provable sample complexity bounds which are almost optimal.

3.2 Separable Random Projections

Recall that from the discussion in Section 2 and the notation introduced in Section 2.1.1, we have the following set of measurements:

$$\begin{aligned}\mathcal{L}_1^{(3)}(\mathbf{X}) &= \begin{bmatrix} \langle A_1 \otimes a, \mathbf{X} \rangle \\ \vdots \\ \langle A_{m_1} \otimes a, \mathbf{X} \rangle \end{bmatrix}, & \mathcal{L}_2^{(3)}(\mathbf{X}) &= \begin{bmatrix} \langle A_1 \otimes b, \mathbf{X} \rangle \\ \vdots \\ \langle A_{m_1} \otimes b, \mathbf{X} \rangle \end{bmatrix}, \\ \mathcal{L}_1^{(1)}(\mathbf{X}) &= \begin{bmatrix} \langle c \otimes B_1, \mathbf{X} \rangle \\ \vdots \\ \langle c \otimes B_{m_2}, \mathbf{X} \rangle \end{bmatrix}, & \mathcal{L}_2^{(1)}(\mathbf{X}) &= \begin{bmatrix} \langle d \otimes B_1, \mathbf{X} \rangle \\ \vdots \\ \langle d \otimes B_{m_2}, \mathbf{X} \rangle \end{bmatrix}.\end{aligned}$$

In the above, each $A_i, B_i \in \mathbb{R}^{n_2 \times n_3}$ is a random Gaussian matrix with i.i.d $\mathcal{N}(0, 1)$ entries, and $a, b \in \mathbb{R}^{n_3}$, $c, d \in \mathbb{R}^{n_1}$ are random vectors distributed uniformly on the unit sphere. Finally, collecting all of the above measurements into a single operator, we have $y = \mathcal{L}(\mathbf{X})$, and the total number of samples is thus $m = 2m_1 + 2m_2$.

In the context of random tensor sensing, (9), (10), (11) and (12) reduce to solving low rank matrix recovery problems from random Gaussian measurements, where the measurements are as detailed in Section 2.1.

The following lemma shows that the observations $\mathcal{L}(\mathbf{X})$ can essentially be thought of as linear Gaussian measurements of the contractions $X_a^3, X_b^3, X_c^1, X_d^1$. This is crucial in reducing the tensor recovery problem to the problem of recovering the tensor contractions, instead.

Lemma 3.1. *For tensor \mathbf{X} , matrix A and vector a of commensurate dimensions,*

$$\langle A \otimes a, \mathbf{X} \rangle = \langle A, X_a^3 \rangle.$$

Similarly, for a vector c and matrix B of commensurate dimensions

$$\langle c \otimes B, \mathbf{X} \rangle = \langle B, X_c^1 \rangle.$$

Proof. We only verify the first equality, the second equality is proved in an identical manner. Let us denote by \mathbf{X}_k the k^{th} mode 3 slice of \mathbf{X} where $k = 1, \dots, n_3$. Then we have,

$$\langle A \otimes a, \mathbf{X} \rangle = \sum_{k=1}^{n_3} a_k \langle A, \mathbf{X}_k \rangle = \langle A, \sum_{k=1}^{n_3} a_k \mathbf{X}_k \rangle = \langle A, X_a^3 \rangle.$$

□

As a consequence of the above lemma, it is easy to see that

$$\langle A \otimes a, \mathbf{X} \rangle = \langle A, X_a^3 \rangle = \langle A, \sum_{i=1}^{n_3} a_i X_i^3 \rangle = \sum_{i=1}^{n_3} a_i \langle A, X_i^3 \rangle,$$

thus establishing separability.

Since X_a^3 and X_b^3 are low-rank matrices, the observation operators $\mathcal{L}_k^{(3)}(\mathbf{X})$ essentially provide Gaussian random projections of X_a^3 and X_b^3 , which in turn can be recovered using matrix-based techniques. The following lemma establishes “faithfulness” in the context of separable random projections.

Lemma 3.2. *Suppose $m_1 > 3r(n_1 + n_2 - r)$. Then the unique solutions to problems (9) and (10) are X_a^3 and X_b^3 respectively with high probability. Similarly, if $m_2 > 3r(n_2 + n_3 - r)$ then the unique solutions to problems (11) and (12) are X_c^1 and X_d^1 respectively with high probability.*

Proof. Again, we only prove the first part of the claim, the second follows in an identical manner. Note that by Lemma 3.1 and Lemma 2.1, X_a^3 and X_b^3 are feasible rank r solutions to (9) and (10) respectively. By Proposition 3.11 of [15], we have that the nuclear norm heuristic succeeds in recovering rank r matrices from $m_1 > 3r(n_1 + n_2 - r)$ with high probability. \square

Remark In this sub-section, we will refer to events which occur with probability exceeding $1 - \exp(-C_0 n_1)$ as events that occur “with high probability” (w.h.p.). We will transparently be able to take appropriate union bounds of high probability events since the number of events being considered is small enough that the union event also holds w.h.p. (thus affecting only the constants involved). Hence, in the subsequent results, we will not need to refer to the precise probabilities.

Since the contractions X_a^3 and X_b^3 of the tensor \mathbf{X} are successfully recovered and the tensor satisfies Assumption 2.1, the second stage of Leurgans’ algorithm can be used to recover the factors u_i and v_i . Similarly, from X_c^1 and X_d^1 , the factors v_i and w_i can be recovered. The above sequence of observations leads to the following sample complexity bound for low rank tensor recovery from random measurements:

Theorem 3.3. *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an unknown tensor of interest with rank $r \leq \min\{n_1, n_2, n_3\}$. Suppose we obtain samples as described by (8). Suppose $m_1 > 3r(n_1 + n_2 - r)$ and $m_2 > 3r(n_2 + n_3 - r)$. Then T-ReCs (Algorithm 2) succeeds in exactly recovering \mathbf{X} and its low rank decomposition (5) with high probability.*

Proof. By Lemma 2.1 $X_a^3, X_b^3, X_c^1, X_d^1$ are all rank at most r . By Lemma 3.1, the tensor observations $y_1^{(3)}, y_2^{(3)}, y_1^{(1)}, y_2^{(1)}$ provide linear Gaussian measurements of $X_a^3, X_b^3, X_c^1, X_d^1$. By Lemma 3.5, the convex problems (9), (10), (11), (12) correctly recover the modal contractions $X_a^3, X_b^3, X_c^1, X_d^1$. Since the vectors a, b, c, d are chosen to be randomly uniformly distributed on the unit sphere, the contractions X_a^3, X_b^3 are non-degenerate and pairwise generic almost surely (and similarly X_c^1, X_d^1). Thus, Lemma 2.2 applies and X_a^3, X_b^3 can be used to correctly recover the factors $u_i, v_i, i = 1, \dots, r$. Again by Lemma 2.2 X_c^1, X_d^1 can be used to correctly recover the factors $v_i, w_i, i = 1, \dots, r$. Note that due to the linear independence of the factors, the linear system of equations involving λ_i is full column rank, over-determined, and has an exact solution. The fact that the result holds with high probability follows because one simply needs to take the union bounds of the probabilities of failure exact recovery of the contractions via the solution of (9), (10), (11), (12). \square

Remarks.

1. Theorem 3.3 yields bounds that are order optimal. Indeed, consider the number of samples $m = 2m_1 + 2m_2 \sim O(r(n_1 + n_2 + n_3))$, which by a counting argument is the same as the number of parameters in an order 3 tensor of rank r .
2. For symmetric tensors with symmetric factorizations of the form $\mathbf{X} = \sum_{l=1}^3 \lambda_l v_l \otimes v_l \otimes v_l$, this method becomes particularly simple. Steps 4, 5, 6 in Algorithm 2 become unnecessary, and the factors are revealed directly in step 3. One then only needs to solve the linear system described in step 7 to recover the scale factors. The sample complexity remains $O(nr)$, nevertheless.

3. Note that for the method we propose, the most computationally expensive step is that of solving low-rank matrix recovery problems where the matrix is of size $n_i \times n_j$ for $i, j = 1, 2, 3$. Fast algorithms with rigorous guarantees exist for solving such problems, and we can use any of these pre-existing methods. An important point to note is that, other methods for minimizing the Tucker rank of a tensor by considering “matricized” tensors solve matrix recovery problems for matrices of size $n_i \times n_j n_k$, which can be far more expensive.
4. Note that the sensing operators $\langle A_i \otimes a, \cdot \rangle$ may seem non-standard (vis-a-vis the compressed sensing literature such as [39]), but are very storage efficient. Indeed, one needs to only store random matrices A_i, B_i and random vectors a, b . Storing each of these operators requires $O(n_1 n_2 + n_3)$ space, and is far more storage efficient than (perhaps the more suggestive) sensing operators of the form $\langle \mathbf{A}_i, \cdot \rangle$, with each \mathbf{A}_i being a random tensor requiring $O(n_1 n_2 n_3)$ space. Similar “low rank” sensing operators have been used for matrix recovery [30, 22].
5. While the results here are presented in the case where the A_i, B_i are random Gaussian matrices and the a, b are uniformly distributed on the sphere, the results are not truly dependent on these distributions. The A_i, B_i need to be structured so that they enable low-rank matrix recovery (i.e., they need to be “faithful”). Hence, for instance it would suffice if the entries of these matrices were sub-Gaussian, or had appropriate restricted isometry properties with respect to low rank matrices [39].

3.3 Tensor Completion

In the context of tensor completion, for a fixed (but unknown) \mathbf{X} , a subset of the entries \mathbf{X}_Ω are revealed for some index set $\Omega \subseteq [n_1] \times [n_2] \times [n_3]$. We assumed that the measurements thus revealed are in a union of four slices. For the i^{th} mode-1 slice let us define

$$\Omega_i^{(1)} := \Omega \cap S_i^{(1)} \quad m_i^{(1)} := |\Omega \cap S_i^{(1)}|.$$

These are precisely the set of entries revealed in the i^{th} mode-1 slice and the corresponding cardinality. Similarly for the k^{th} mode-3 slice we define

$$\Omega_k^{(3)} = \Omega \cap S_k^{(3)} \quad m_k^{(3)} := |\Omega \cap S_k^{(3)}|. \quad (13)$$

We will require the existence of two distinct mode-1 slices (say i_1^* and i_2^*) from which measurements are obtained. Indeed,

$$\mathcal{L}_1^{(1)}(\mathbf{X}) := (\mathbf{X})_{\Omega_{i_1^*}^{(1)}} \quad \mathcal{L}_2^{(1)}(\mathbf{X}) := (\mathbf{X})_{\Omega_{i_2^*}^{(1)}}. \quad (14)$$

Similarly we will also require the existence of two different slices in mode 3 ⁸ (say k_1^* and k_2^*) from which we have measurements:

$$\mathcal{L}_1^{(3)}(\mathbf{X}) := (\mathbf{X})_{\Omega_{k_1^*}^{(3)}} \quad \mathcal{L}_2^{(3)}(\mathbf{X}) := (\mathbf{X})_{\Omega_{k_2^*}^{(3)}}.$$

We will require the cardinalities of the measurements from mode 1, $m_{i_1^*}^{(1)}$ and $m_{i_2^*}^{(1)}$ and from mode 3, $m_{k_1^*}^{(3)}$ and $m_{k_2^*}^{(3)}$ to be sufficiently large so that they are faithful (to be made precise subsequently),

⁸We choose modes 1 and 3 arbitrarily. Any two of the 3 modes suffice.

and this will determine the sample complexity. The key aspect of the algorithm is that it *only makes use of the samples in these four distinct slices*. No other samples outside these four slices need be revealed at all (so that all the other $m_i^{(1)}$ and $m_k^{(3)}$ can be zero). The indices sampled from each slice are drawn uniformly and randomly without replacement. Note that for a specified $m_{i_1^*}^{(1)}$, $m_{i_2^*}^{(1)}$, $m_{k_1^*}^{(3)}$ and $m_{k_2^*}^{(3)}$ the overall sample complexity implied is $m_{i_1^*}^{(1)} + m_{i_2^*}^{(1)} + m_{k_1^*}^{(3)} + m_{k_2^*}^{(3)}$.

In the context of tensor completion, (9), (10), (11) and (12) reduce to solving low rank matrix completion problems for the slices $S_{i_1^*}^{(1)}$, $S_{i_2^*}^{(1)}$, $S_{k_1^*}^{(3)}$, $S_{k_2^*}^{(3)}$. Contraction recovery in this context amounts to obtaining complete slices, which can then be used as inputs to Leurgans' algorithm. There are a few important differences however, when compared to the case of recovery from Gaussian random projections. For the matrix completion sub-steps to succeed, we need the following standard incoherence assumptions from the matrix completion literature [38].

Let \mathcal{U}, \mathcal{V} and \mathcal{W} represent the linear spans of the vectors $\{u_i\}_{i=1,\dots,r}$, $\{v_i\}_{i=1,\dots,r}$, $\{w_i\}_{i=1,\dots,r}$. Let $P_{\mathcal{U}}$, $P_{\mathcal{V}}$ and $P_{\mathcal{W}}$ respectively represent the projection operators corresponding to \mathcal{U} , \mathcal{V} and \mathcal{W} . The coherence of the subspace \mathcal{U} (similarly for \mathcal{V} and \mathcal{W}) is defined as:

$$\mu(\mathcal{U}) := \frac{n_1}{r} \max_{i=1,\dots,n_1} \|P_{\mathcal{U}}(e_i)\|^2,$$

where $\{e_i\}$ are the canonical basis vectors.

Assumption 3.1 (Incoherence). $\mu_0 := \max\{\mu(\mathcal{U}), \mu(\mathcal{V}), \mu(\mathcal{W})\}$ is a positive constant independent of the rank and the dimensions of the tensor.

Such an incoherence condition is required in order to be able to complete the matrix slices from the observed data [38]. We will see subsequently that when the tensor is of rank r , so are the different slices of the tensor and each slice will have a “thin” singular value decomposition. Furthermore, the incoherence assumption will also hold for these slices.

Definition 3.1. Let $X_i^1 = U\Sigma V^T$ be the singular value decomposition of the tensor slice X_i^1 . We say that the tensor \mathbf{X} satisfies the slice condition for slice $S_i^{(1)}$ with constant $\mu_i^{(1)}$ if the element-wise infinity (max) norm

$$\|UV^T\|_{\infty} \leq \mu_i^{(1)} \sqrt{\frac{r}{n_2 n_3}}.$$

The slice condition is analogously defined for the slices along other modes, i.e. $S_j^{(2)}$ and $S_k^{(3)}$. We will denote by $\mu_j^{(2)}$ and $\mu_k^{(3)}$ the corresponding slice constants. We will require our distinct slices from which samples are obtained to satisfy these slice conditions.

Remark The slice conditions are standard in the matrix completion literature, see for instance [38]. As pointed out in [38], the slice conditions are not much more restrictive than the incoherence condition, because if the incoherence condition is satisfied with constant μ_0 then (by a simple application of the Cauchy-Schwartz inequality) the slice condition for $S_i^{(1)}$ is also satisfied with constant $\mu_1(i) \leq \mu_0 \sqrt{r}$ for all i (and similarly for $\mu_j^{(2)}$ and $\mu_k^{(3)}$). Hence, the slice conditions can be done away with, and using this weaker bound only increases the sample complexity bound for exact reconstruction by a multiplicative factor of r .

Remark Note that the incoherence assumption and the slice condition are known to be satisfied for suitable random ensembles of models, such as the random orthogonal model, and models where the singular vectors are bounded element-wise [38].

The decomposition (5) ties factor information about the tensor to factor information of contractions. A direct corollary of Lemma 2.1 is that contraction matrices are incoherent whenever the tensor is incoherent:

Corollary 3.4. *If the tensor satisfies the incoherence assumption, then so do the contractions. Specifically all the tensor slices satisfy incoherence.*

Proof. Consider for instance the slices X_k^3 for $k = 1, \dots, n_3$. By Lemma 2.1, the row and column-spaces of each slice are precisely \mathcal{U} and \mathcal{V} respectively, thus the incoherence assumption also holds for the slices. \square

We now detail our result for the tensor completion problem:

Lemma 3.5. *Given a tensor \mathbf{X} with rank $r \leq n_1$ which satisfies the following:*

- *Assumptions 2.2 and 3.1,*
- *The samples are obtained as described in (13), (14).*
- *Suppose the number of samples from each slice satisfy:*

$$\begin{aligned} m_{i_1^*}^{(1)} &\geq 32 \max \left\{ \mu_0, \left(\mu_{i_1^*}^{(1)} \right)^2 \right\} r (n_2 + n_3) \log^2 n_3 \\ m_{i_2^*}^{(1)} &\geq 32 \max \left\{ \mu_0, \left(\mu_{i_2^*}^{(1)} \right)^2 \right\} r (n_2 + n_3) \log^2 n_3 \\ m_{k_1^*}^{(3)} &\geq 32 \max \left\{ \mu_0, \left(\mu_{k_1^*}^{(3)} \right)^2 \right\} r (n_1 + n_2) \log^2 n_2 \\ m_{k_2^*}^{(3)} &\geq 32 \max \left\{ \mu_0, \left(\mu_{k_2^*}^{(3)} \right)^2 \right\} r (n_1 + n_2) \log^2 n_2 \end{aligned}$$

- *The slice condition (Definition 3.1) for each of the four slices $S_{i_1^*}^{(1)}, S_{i_2^*}^{(1)}, S_{k_1^*}^{(3)}, S_{k_2^*}^{(3)}$ hold.*

Then the unique solutions to problems (9), (10), (11) and (12) are $X_{k_1^}^3, X_{k_2^*}^3, X_{i_1^*}^1$, and $X_{i_2^*}^1$ respectively with probability exceeding $1 - C \log(n_2) n_2^{-\beta}$ for some constants $C, \beta > 0$.*

Proof. By Lemma 2.1 $X_{i_1^*}^1, X_{i_2^*}^1, X_{k_1^*}^3, X_{k_2^*}^3$ are all rank at most r . By Theorem 1.1 of [38], the convex problems (9), (10), (11), (12) correctly recover the full slices $X_{k_1^*}^3, X_{k_2^*}^3, X_{i_1^*}^1, X_{i_2^*}^1$ with high probability. (Note that the relevant incoherence conditions in [38] are satisfied due to Corollary 2.1 and the slice condition assumption. Furthermore the number of samples specified meets the sample complexity requirements of Theorem 1.1 in [38] for exact recovery.) \square

Remark We note that in this sub-section, events that occur with probability exceeding $1 - C \log(n_2) n_2^{-\beta}$ (recall that $n_1 \leq n_2 \leq n_3$) are termed as occurring with high probability (w.h.p.). We will transparently be able to union bound these events (thus changing only the constants) and hence we refrain from mentioning these probabilities explicitly.

Theorem 3.6. *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an unknown tensor of interest with rank $r \leq n_1$, such that the tensor slices $X_{k_1^*}^3, X_{k_2^*}^3$ are non-degenerate and pairwise generic, and similarly $X_{i_1^*}^1, X_{i_2^*}^1$ are non-degenerate and pairwise generic. Then, under the same set of assumptions made for Lemma 3.5, the procedure outlined in Algorithm 2 succeeds in exactly recovering \mathbf{X} and its low rank decomposition (1) with high probability.*

Proof. The proof follows along the same lines as that of Theorem 3.3, with Lemma 3.5 allowing us to exactly recover the slices $X_{i_1^*}^1, X_{i_2^*}^1, X_{k_1^*}^3, X_{k_2^*}^3$. Since these slices satisfy non-degeneracy and pairwise genericity, the tensor factors $u_i, v_i, w_i, i = 1, \dots, r$ can be exactly recovered (up to scaling) by following steps (3), (5) and (6) of Algorithm 2. Also, the system of equations to recover λ is given by

$$\mathbf{X}_\Omega = \sum_{i=1}^r \lambda_i (u_i \otimes v_i \otimes w_i)_\Omega.$$

□

Remarks.

1. Theorem 3.6 yields bounds that are almost order optimal when μ_0 and $\mu_k^{(i)}$ are constant (independent of r and the dimension). Indeed, the total number of samples required is $m \sim O(r n_3 \log^2 n_3)$, which by a counting argument is nearly the same number of parameters in an order 3 tensor of rank r (except for the additional logarithmic factor).
2. The comments about efficiency for symmetric factorizations in the Gaussian random projections case hold here as well.
3. We do not necessarily need sampling without replacement from the four slices. Similar results can be obtained for other sampling models such as with replacement [38], and even non-uniform sampling [17]. Furthermore, while the method proposed here for the task of matrix completion relies on nuclear norm minimization, a number of other approaches such as alternating minimization [27, 17, 8] can also be adopted; our algorithm relies only on the successful completion of the slices.
4. Note that we can remove the slice condition altogether since the incoherence assumption implies the slice condition with $\mu_1 = \mu_0 \sqrt{r}$. Removing the slice condition then implies an overall sample complexity of $O(r^2 n_3 \log^2 n_3)$.

4 Extension to Higher Order Tensors

The results of Section 3 can be extended to higher order tensors in a straightforward way. While the ideas remain essentially the same, the notation is necessarily more cumbersome in this section. We omit some technical proofs to avoid repetition of closely analogous arguments from the third order case, and focus on illustrating how to extend the methods to the higher order setting.

Consider a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$ of order K and dimension $n_1 \times \dots \times n_K$. Let us assume, without loss of generality, that $n_1 \leq n_2 \leq \dots \leq n_K$. Let the rank of this tensor be $r \leq n_1$ and be

given by the decomposition:

$$\mathbf{X} = \sum_{l=1}^r u_l^1 \otimes \cdots \otimes u_l^K = \sum_{l=1}^r \bigotimes_{p=1}^K u_l^p,$$

where $u_l^p \in \mathbb{R}^{n_p}$. We will be interested in slices of the given tensor that are identified by picking two *consecutive* modes $(k, k+1)$, and by fixing all the indices not in those modes, i.e. $i_1 \in [n_1], \dots, i_{k-1} \in [n_{k-1}], i_{k+2} \in [n_{k+2}], \dots, i_K \in [n_K]$. Thus the indices of a slice S are:

$$S := \{i_1\} \times \cdots \times \{i_{k-1}\} \times [n_k] \times [n_{k+1}] \times \{i_{k+2}\} \times \cdots \times \{i_K\},$$

and the corresponding slice may be viewed as a matrix, denoted by \mathbf{X}_S . While slices of tensors can be defined more generally (i.e. the modes need not be consecutive), in this paper we will only need to deal with such “contiguous” slices.⁹ We will denote the collection of all slices where modes $(k, k+1)$ are contained to be:

$$\mathcal{S}^{(k)} := \{\{i_1\} \times \cdots \times \{i_{k-1}\} \times [n_k] \times [n_{k+1}] \times \{i_{k+2}\} \times \cdots \times \{i_K\} \mid i_1 \in [n_1], \dots, i_K \in [n_K]\}.$$

Every element of $\mathcal{S}^{(k)}$ is a set of indices, and we can identify a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times \cdots \times n_{k-1} \times n_{k+2} \times \cdots \times n_K}$ with a map $\mathcal{A} : \mathcal{S}^{(k)} \rightarrow \mathbb{R}$. Using this identification, every element of \mathbf{A} can thus also be referenced by $S \in \mathcal{S}^{(k)}$. To keep our notation succinct, we will thus refer to \mathbf{A}_S as the element corresponding to S under this identification. Thus if $S = \{i_1\} \times \cdots \times \{i_{k-1}\} \times [n_k] \times [n_{k+1}] \times \{i_{k+2}\} \times \cdots \times \{i_K\}$, the element:

$$\mathbf{A}_S = \mathbf{A}_{i_1, \dots, i_{k-1}, i_{k+2}, \dots, i_K}.$$

Using this notation, we can define a high-order contraction. A mode- k contraction of \mathbf{X} with respect to a tensor \mathbf{A} is thus:

$$X_{\mathbf{A}}^k := \sum_{S \in \mathcal{S}^{(k)}} \mathbf{A}_S \mathbf{X}_S. \quad (15)$$

Note that since $X_{\mathbf{A}}^k$ is a sum of (two-dimensional) slices, it is a matrix. As in the third order case, we will be interested in contractions where \mathbf{A} is either random or a coordinate tensor. The analogue of Lemma 2.2 for the higher order case is the following:

Lemma 4.1. *Let \mathbf{X} have the decomposition $\mathbf{X} = \sum_{l=1}^r \bigotimes_{p=1}^K u_l^p$. Then we have that the contraction $X_{\mathbf{A}}^k$ has the following matrix decomposition:*

$$X_{\mathbf{A}}^k = \sum_{l=1}^r \nu_l^k u_l^k \left(u_l^{k+1} \right)^T, \quad (16)$$

where $\nu_l^k := \langle \mathbf{A}, \bigotimes_{p \neq k, k+1} u_l^p \rangle$. Furthermore, if $X_{\mathbf{B}}^k$ is another contraction with respect to \mathbf{B} , then the eigenvectors of the matrices

$$M_1 = X_{\mathbf{A}}^k \left(X_{\mathbf{B}}^k \right)^\dagger \quad M_2 = \left(\left(X_{\mathbf{B}}^k \right)^\dagger X_{\mathbf{A}}^k \right)^T \quad (17)$$

respectively are $\{u_l^k\}_{l=1, \dots, r}$ and $\{u_l^{k+1}\}_{l=1, \dots, r}$.

⁹In general, a slice corresponding to any pair of modes (k_1, k_2) suffices for our approach. However, to keep the notation simple we present the case where slices correspond to mode pairs of the form $(k, k+1)$.

Proof. It is straightforward to verify by simply expanding the definition of $X_{\mathbf{A}}^k$ using the definition of contraction (15):

$$\left[X_{\mathbf{A}}^k\right]_{j_k, j_{k+1}} = \sum_{j_1, \dots, j_{k-1}, j_{k+2}, \dots, j_K} \sum_{l=1}^r \left(\prod_{p=1}^K (u_l^p)_{j_k} \right) \mathbf{A}_{j_1, \dots, j_{k-1}, j_{k+2}, \dots, j_K}.$$

Rearranging terms, we get the decomposition (16). The eigenvalues of M_1, M_2 follow along similar lines to the proof of Lemma 2.2. \square

As a consequence of the above lemma, if \mathbf{X} is of low rank, so are all the contractions. The notions of non-degeneracy and pairwise genericity of contractions extend in a natural way to the higher order case. We say that a contraction $X_{\mathbf{A}}^k$ is non-degenerate if $\nu_l^k \neq 0$ for all $l = 1, \dots, r$. Furthermore, a pair of contractions is pairwise generic if the corresponding ratios ν_l^k are all distinct for $l = 1, \dots, r$. Non-degeneracy and pairwise genericity hold almost surely when the contractions are computed with random tensors \mathbf{A}, \mathbf{B} from appropriate random ensembles (e.g. *i.i.d.* normally distributed entries). In much the same way as the third order case, Leurgans' algorithm can be used to perform decomposition of low-rank tensors using Lemma 4.1. This is described in Algorithm 3.

Algorithm 3 Leurgans' Algorithm for Higher Order Tensors

- 1: **Input:** Tensor \mathbf{X} .
- 2: **for** $k = 1$ to $K - 1$ **do**
- 3: Compute contractions $X_{\mathbf{A}}^k$ and $X_{\mathbf{B}}^k$ for some tensors \mathbf{A} and \mathbf{B} of appropriate dimensions, such that the contractions are non-degenerate and pairwise generic.
- 4: Compute eigen-decompositions of $M_1 := X_{\mathbf{A}}^k (X_{\mathbf{B}}^k)^\dagger$ and $M_2 := (X_{\mathbf{B}}^k)^\dagger X_{\mathbf{A}}^k$. Let \tilde{U}^k and \tilde{U}^{k+1} denote the matrices whose columns are the eigenvectors of M_1 and M_2^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_1 and M_2 .)
- 5: If $k = 1$, let $U^1 := \tilde{U}^1$ and $U^2 := \tilde{U}^2$.
- 6: If $k \geq 2$, simultaneously reorder the columns of $\tilde{U}^k, \tilde{U}^{k+1}$, also performing simultaneous sign reversals as necessary so that the columns of \tilde{U}^k obtained match with the columns of U^k (obtained in the previous iteration), call the resulting matrices U^k, U^{k+1} . (The eigenvectors corresponding to mode $k + 1$, thus obtained are denoted as $\{u_l^{k+1}\}_{l=1, \dots, r}$.)
- 7: **end for**
- 8: Solve for λ_l in the (over-determined) linear system

$$\mathbf{X} = \sum_{l=1}^r \lambda_l \bigotimes_{k=1}^K u_l^k.$$

- 9: **Output:** Recovered tensor $\mathbf{X} = \sum_{l=1}^r \lambda_l \bigotimes_{k=1}^K u_l^k$.
-

Finally, the notion of *separable measurements* can be extended to higher order tensors in a natural way.

Definition 4.1. Consider a linear operator $\mathcal{L} : \mathbb{R}^{n_1 \times \dots \times n_K} \rightarrow \mathbb{R}^n$. We say that \mathcal{L} is separable with respect to the k^{th} mode if there exist $\mathbf{W} \in \mathbb{R}^{n_1 \times \dots \times n_{k-1} \times n_{k+2} \times \dots \times n_K}$ and a linear operator

$\mathcal{T}^{(k)} : \mathbb{R}^{n_k \times n_{k+1}} \rightarrow \mathbb{R}^n$, such that for every $\mathbf{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$:

$$\mathcal{L}(\mathbf{X}) = \sum_{S \in \mathcal{S}^{(k)}} \mathbf{W}_S \mathcal{T}^{(k)}(X_S^k).$$

Analogous to the third order case, we assume that we are presented with two sets of separable measurements per mode:

$$\begin{aligned} y_1^{(k)} &= \mathcal{L}_1^{(k)}(\mathbf{X}) = \sum_{S \in \mathcal{S}^{(k)}} (\mathbf{W}_1)_S \mathcal{T}^{(k)}(X_S^k) \\ y_2^{(k)} &= \mathcal{L}_2^{(k)}(\mathbf{X}) = \sum_{S \in \mathcal{S}^{(k)}} (\mathbf{W}_2)_S \mathcal{T}^{(k)}(X_S^k) \end{aligned}$$

for $k = 1, \dots, K-1$ with each of $y_1^{(k)}, y_2^{(k)} \in \mathbb{R}^{m_k}$. Once again, by separability we have:

$$y_1^{(k)} = \mathcal{T}^{(k)}(X_{\mathbf{W}_1}^k) \quad y_2^{(k)} = \mathcal{T}^{(k)}(X_{\mathbf{W}_2}^k),$$

and since the contractions $X_{\mathbf{W}_1}^k$ and $X_{\mathbf{W}_2}^k$ are low rank, nuclear norm minimization can be used to recover these contractions via:

$$\underset{Z_1}{\text{minimize}} \quad \|Z_1\|_* \quad \text{subject to} \quad y_1^{(k)} = \mathcal{T}^{(k)}(Z_1), \quad (18)$$

$$\underset{Z_2}{\text{minimize}} \quad \|Z_2\|_* \quad \text{subject to} \quad y_2^{(k)} = \mathcal{T}^{(k)}(Z_2), \quad (19)$$

for each $k = 1, \dots, K-1$. After recovering the two contractions for each mode, we can then apply (the higher order) Leurgans' algorithm to recover the tensor factors. The precise algorithm is described in Algorithm 4. Provided the $\mathcal{T}^{(k)}(\cdot)$ are faithful, the tensor contractions can be successfully recovered via nuclear norm minimization. Furthermore, if the contractions are non-degenerate and pairwise generic, the method can successfully recover the entire tensor.

4.1 Separable Random Projections

Given tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \dots \times n_{K_1}}$, $\mathbf{B} \in \mathbb{R}^{n_{K_1+1} \times \dots \times n_{K_1+K_2}}$, $\mathbf{C} \in \mathbb{R}^{n_{K_1+K_2+1} \times \dots \times n_{K_1+K_2+K_3}}$ of orders K_1 , K_2 and K_3 respectively with $K_1 + K_2 + K_3 = K$, we define their outer product as:

$$[\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}]_{i_1, \dots, i_K} := [\mathbf{A}]_{i_1, \dots, i_{K_1}} [\mathbf{B}]_{i_{K_1+1}, \dots, i_{K_1+K_2}} [\mathbf{C}]_{i_{K_1+K_2+1}, \dots, i_{K_1+K_2+K_3}}$$

Note also that the inner-product for higher order tensors is defined in the natural way:

$$\langle \mathbf{T}, \mathbf{X} \rangle := \sum_{i_1, \dots, i_K} [\mathbf{T}]_{i_1, \dots, i_K} [\mathbf{X}]_{i_1, \dots, i_K}.$$

In this higher order setting, we also work with specific separable random projection operators,

Algorithm 4 T-ReCs for Higher Order Tensors

- 1: **Input:** Measurements $y_i^{(k)} = \mathcal{L}_i^{(k)}(\mathbf{X})$, for $k = 1, \dots, K$, $i = 1, 2$.
- 2: **for** $k = 1$ to $K - 1$ **do**
- 3: Solve convex optimization problems (18) and (19) to obtain optimal solutions Z_1^* and Z_2^* respectively.
- 4: Compute eigen-decompositions of $M_1 := Z_1^*(Z_2^*)^\dagger$ and $M_2 := (Z_2^*)^\dagger Z_1^*$. Let \tilde{U}^k and \tilde{U}^{k+1} denote the matrices whose columns are the normalized eigenvectors of M_1 and M_2^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_1 and M_2^T .)
- 5: If $k = 1$, let $U^1 := \tilde{U}^1$ and $U^2 := \tilde{U}^2$.
- 6: If $k \geq 2$, simultaneously reorder the columns of $\tilde{U}^k, \tilde{U}^{k+1}$, also performing simultaneous sign reversals as necessary so that the columns of \tilde{U}^k obtained match with the columns of U^k (obtained in the previous iteration), call the resulting matrices U^k, U^{k+1} . (The eigenvectors corresponding to mode $k + 1$, thus obtained are denoted as $\{u_l^{k+1}\}_{l=1, \dots, r}$.)
- 7: **end for**
- 8: Solve for λ_l in the (over-determined) linear system

$$y_i^{(k)} = \sum_{l=1}^r \lambda_l \mathcal{L}_i^{(k)} \left(\bigotimes_{k=1}^K u_l^k \right), \quad k = 1, \dots, K - 1, \quad i = 1, 2.$$

- 9: **Output:** Recovered tensor $\mathbf{X} = \sum_{l=1}^r \lambda_l \bigotimes_{k=1}^K u_l^k$.
-

which are defined as below:

$$y_1^{(k)} = \mathcal{L}_1^{(k)}(\mathbf{X}) := \begin{bmatrix} \langle \mathbf{A}_k \otimes \Gamma_1^{(k)} \otimes \mathbf{B}_k, \mathbf{X} \rangle \\ \vdots \\ \langle \mathbf{A}_k \otimes \Gamma_{m_k}^{(k)} \otimes \mathbf{B}_k, \mathbf{X} \rangle \end{bmatrix} \quad (20)$$

$$y_2^{(k)} = \mathcal{L}_2^{(k)}(\mathbf{X}) := \begin{bmatrix} \langle \mathbf{C}_k \otimes \Gamma_1^{(k)} \otimes \mathbf{D}_k, \mathbf{X} \rangle \\ \vdots \\ \langle \mathbf{C}_k \otimes \Gamma_{m_k}^{(k)} \otimes \mathbf{D}_k, \mathbf{X} \rangle \end{bmatrix} \quad (21)$$

In the above expressions, $\mathbf{A}_k, \mathbf{C}_k \in \mathbb{R}^{n_1 \times \dots \times n_{k-1}}$, and $\mathbf{B}_k, \mathbf{D}_k \in \mathbb{R}^{n_{k+2} \times \dots \times n_K}$. The tensors $\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k$ are all chosen so that their entries are randomly and independently distributed according to $\mathcal{N}(0, 1)$ and subsequently normalized to have unit Euclidean norm. The matrices $\Gamma_i^{(k)} \in \mathbb{R}^{n_k \times n_{k+1}}$ for $i = 1, \dots, m_k$ have entries randomly and independently distributed according to $\mathcal{N}(0, 1)$. For each k we have $2m_k$ measurements so that in total there are $2 \sum_{i=1}^{K-1} m_k$ measurements.

Lemma 4.2. *We have the following identity:*

$$\langle \mathbf{A}_k \otimes \Gamma_i^{(k)} \otimes \mathbf{B}_k, \mathbf{X} \rangle = \langle \Gamma_i^{(k)}, X_{\mathbf{A}_k \otimes \mathbf{B}_k}^k \rangle.$$

Proof. The proof is analogous to that of Lemma 3.1.

$$\begin{aligned}
& \langle \mathbf{A}_k \otimes \Gamma_i^{(k)} \otimes \mathbf{B}_k, \mathbf{X} \rangle \\
&= \sum_{l=1}^r \langle \mathbf{A}_k \otimes \Gamma_i^{(k)} \otimes \mathbf{B}_k, \bigotimes_{p=1}^K u_l^p \rangle \\
&\stackrel{(i)}{=} \sum_{l=1}^r \langle \mathbf{A}_k, \bigotimes_{p=1}^{k-1} u_l^p \rangle \langle \mathbf{B}_k, \bigotimes_{p=k+2}^K u_l^p \rangle \langle \Gamma_i^{(k)}, u_l^k \otimes u_l^{k+1} \rangle \\
&= \sum_{l=1}^r \langle \mathbf{A}_k \otimes \mathbf{B}_k, \bigotimes_{p=1}^{k-1} u_l^p \bigotimes_{p=k+2}^K u_l^p \rangle \langle \Gamma_i^{(k)}, u_l^k \otimes u_l^{k+1} \rangle \\
&\stackrel{(ii)}{=} \sum_{l=1}^r \nu_l^k \langle \Gamma_i^{(k)}, u_l^k \otimes u_l^{k+1} \rangle \quad (\text{where } \nu_l^k = \langle \mathbf{A}_k \otimes \mathbf{B}_k, \bigotimes_{p \neq k, k+1}^K u_l^p \rangle) \\
&= \langle \Gamma_i^{(k)}, \sum_{l=1}^r \nu_l^k u_l^k \otimes u_l^{k+1} \rangle \\
&= \langle \Gamma_i^{(k)}, X_{\mathbf{A}_k \otimes \mathbf{B}_k}^k \rangle.
\end{aligned}$$

The equality (i) follows from the identity $\langle a \otimes b \otimes c, x \otimes y \otimes z \rangle = \langle a, x \rangle \langle b, y \rangle \langle c, z \rangle$ for a, b, c, x, y, z of commensurate dimensions. The equality (ii) follows from the definition ν_k^l in Lemma 4.1 \square

It follows immediately from Lemma 4.2 that in (21), for each $k = 1, \dots, K-1$, $\mathcal{L}_1^{(k)}(\cdot)$, $\mathcal{L}_2^{(k)}(\cdot)$ are in fact, separable so that Algorithm 4 is applicable. Recovering the contractions involves solving a set of nuclear norm minimization sub-problems for each $k = 1, \dots, K-1$:

$$\begin{aligned}
& \underset{Z_1}{\text{minimize}} && \|Z_1\|_* \\
& \text{subject to} && y_1^{(k)} = \begin{bmatrix} \langle \Gamma_1^{(k)}, Z_1 \rangle \\ \vdots \\ \langle \Gamma_{m_k}^{(k)}, Z_1 \rangle \end{bmatrix}
\end{aligned} \tag{22}$$

$$\begin{aligned}
& \underset{Z_2}{\text{minimize}} && \|Z_2\|_* \\
& \text{subject to} && y_2^{(k)} = \begin{bmatrix} \langle \Gamma_1^{(k)}, Z_2 \rangle \\ \vdots \\ \langle \Gamma_{m_k}^{(k)}, Z_2 \rangle \end{bmatrix}
\end{aligned} \tag{23}$$

We have the following lemma concerning the solutions of these optimization problems:

Lemma 4.3. *Suppose $m_k > 3r(n_k + n_{k+1} - r)$. Then the unique solutions to problems (22) and (23) are $X_{\mathbf{A}_k \otimes \mathbf{B}_k}^k$ and $X_{\mathbf{C}_k \otimes \mathbf{D}_k}^k$ respectively with high probability.*

The proof is analogous to that of Lemma 3.5.

We have the following theorem concerning the performance of Algorithm 4.

Theorem 4.4. *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$ be an unknown tensor of interest with rank $r \leq \min \{n_1, \dots, n_K\}$. Suppose $m_k > 3r(n_k + n_{k+1} - r)$ for each $k = 1, \dots, K-1$. Then the procedure outlined in Algorithm 4 succeeds in exactly recovering \mathbf{X} and its low rank decomposition with high probability.*

The proof parallels that of the proof of Theorem 3.3 and is omitted for the sake of brevity.

Remarks.

- Note that the overall sample complexity is $2 \sum_{k=1}^K m_k$, i.e., $6 \sum_{k=1}^{K-1} r(n_k + n_{k+1} - r)$. This constitutes an order optimal sample complexity because a tensor of rank r and order K of these dimensions has $r \sum_{k=1}^K n_k$ degrees of freedom. In particular, when the tensor is “square” i.e., $n_1 = \dots = n_K = n$ the number of degrees of freedom is Knr whereas the achieved sample complexity is no larger than $12Knr$, i.e., $O(Knr)$.
- As with the third order case, the algorithm is tractable. The main operations involve solving a set of matrix nuclear norm minimization problems (i.e. convex programs), computing eigenvectors, and aligning them. All of these are routine, efficiently solvable steps (thus “polynomial time”) and indeed enables our algorithm to be scalable.

4.2 Tensor Completion

The method described in Section 3 for tensor completion can also be extended to higher order tensors in a straightforward way. Consider a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$ of order K and dimensions $n_1 \times \dots \times n_K$. Let the rank of this tensor be $r \leq \min \{n_1, \dots, n_K\}$ and be given by the decomposition:

$$\mathbf{X} = \sum_{l=1}^r u_l^1 \otimes \dots \otimes u_l^K = \sum_{l=1}^r \bigotimes_{p=1}^K u_l^p,$$

where $u_l^p \in \mathbb{R}^{n_p}$. Extending the sampling notation for tensor completion from the third order case, we define Ω to be the set of indices corresponding to the observed entries of the unknown low rank tensor \mathbf{X} , and define:

$$\Omega^{(k)} := S^{(k)} \cap \Omega \quad m^{(k)} := |\Omega^{(k)}|,$$

where $S^{(k)} \in \mathcal{S}^k$. Akin to the third-order case, along each pair of consecutive modes, we will need samples from two distinguished slices. We denote the index set of these distinct slices by $S_1^{(k)}$ and $S_2^{(k)}$, the corresponding slices by X_1^k and X_2^k , the index set of the samples revealed from these slices by $\Omega_1^{(k)}$ and $\Omega_2^{(k)}$, and their cardinality by $m_1^{(k)}$ and $m_2^{(k)}$.

It is a straightforward exercise to argue that observations obtained from each slice $S_i^{(k)}$, $i = 1, 2$ correspond to separable measurements, so that Algorithm 4 applies. The first step of Algorithm 4 involves solving a set of nuclear norm minimization sub-problems (two problems for each $k = 1, \dots, K-1$) to recover the slices:

$$\underset{Z_1}{\text{minimize}} \quad \|Z_1\|_* \quad \text{subject to} \quad \mathbf{X}_{\Omega_1^{(k)}} = [Z_1]_{\Omega_1^{(k)}} \quad (24)$$

$$\underset{Z_2}{\text{minimize}} \quad \|Z_2\|_* \quad \text{subject to} \quad \mathbf{X}_{\Omega_2^{(k)}} = [Z_2]_{\Omega_2^{(k)}} \quad (25)$$

Each of these optimization problems will succeed in recovering the corresponding slices provided incoherence, non-degeneracy and the slice conditions hold (note that these notions all extend to the higher order setting in a transparent manner). Under these assumptions, if the entries from each slice are uniformly randomly sampled with cardinality at least $m_k^{(i)} > C(n_k + n_{k+1}) \log^2(n_{k+1})$, $i = 1, 2$ for some constant C , then the unique solutions to problems (22) and (23) will be X_1^k and X_2^k respectively with high probability.

Once the slices X_1^k and X_2^k are recovered correctly using (22), (23) for each $k = 1, \dots, K-1$ one can compute $M_1 := X_1^k (X_2^k)^\dagger$ and $M_2 := (X_2^k)^\dagger X_1^k$ and perform eigen-decompositions to obtain the factors (up to possible rescaling) $\{u_l^k\}$ and $\{u_l^{k+1}\}$. Finally, once the tensor factors are recovered, the tensor itself can be recovered exactly by solving a system of linear equations. These observations can be summarized by the following theorem:

Theorem 4.5. *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$ be an unknown tensor of interest with rank $r \leq \min\{n_1, \dots, n_K\}$. Suppose we obtain $m_1^{(k)}$ and $m_2^{(k)}$ random samples from each of the two distinct mode k slices for each $k = 1, \dots, K-1$. Furthermore suppose the tensor \mathbf{X} is incoherent, satisfies the slice conditions for each mode, and the slices from which samples are obtained satisfy non-degeneracy and pairwise genericity for each mode. Then there exists a constant C such that if*

$$m_i^{(k)} > C(n_k + n_{k+1}) \log^2(n_{k+1}) \quad i \in \{1, 2\}$$

the procedure outlined in Algorithm 4 succeeds in exactly recovering \mathbf{X} and its low rank decomposition with high probability.

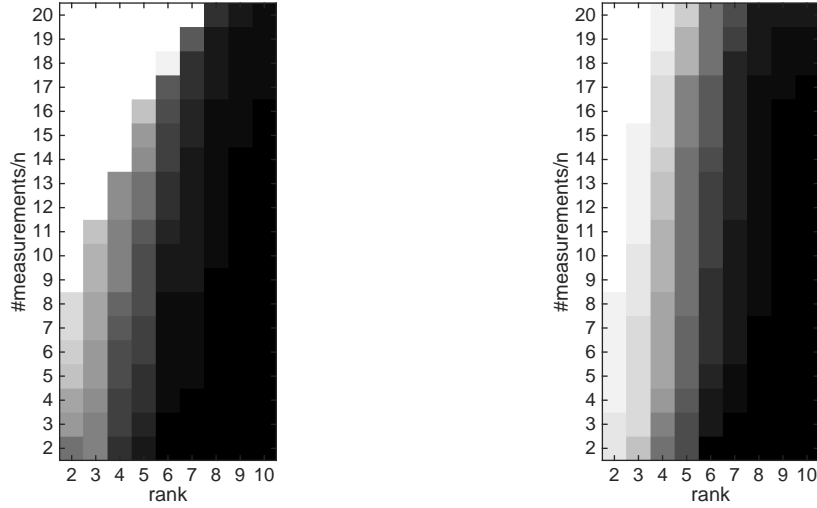
We finally remark that the resulting sample complexity of the entire algorithm is $\sum_{k=1}^{K-1} (m_1^{(k)} + m_2^{(k)})$, which is $O(Krn_K \log^2(n_K))$.

5 Experiments

In this section we present numerical evidence in support of our algorithm. We conduct experiments involving (suitably) random low-rank target tensors and their recovery from (a) Separable Random Projections and (b) Tensor Completion. We obtain phase transition plots for the same, and compare our performance to that obtained from the matrix-unfolding based approach proposed in [35]. For the phase transition plots, we implemented matrix completion using the method proposed in [37], since the SDP approach for exact matrix completion of unfolded tensors was found to be impractical for even moderate-sized problems.

5.1 Separable Random Projections : Phase Transition

In this section, we run experiments comparing T-ReCs to tensor recovery methods based on “matricizing” the tensor via unfolding [35]. We consider a tensor of size $30 \times 30 \times 30$ whose factors $U, V, W \in \mathbb{R}^{n \times r}$ are i.i.d standard Gaussian entries. We vary the rank r from 2 to 10, and look to recover these tensors from different number of measurements $m \in [2, 20] * n$. For each (r, n) pair, we repeated the experiment 10 times, and consider recovery a “success” if the MSE is less than 10^{-5} . Figure 1 shows that the number of measurements needed for accurate tensor recovery is typically less in our method, compared to the ones where the entire tensor is converted to a matrix for low rank recovery.



(a) Tensor recovery using T-ReCs. (n=30)

(b) Tensor recovery using [35]. (n=30)

Figure 1: Phase transition diagram for tensor recovery using our method. White indicates a probability of recovery of 1, while black indicates failure of exact recovery. Note that in the matrix unfolding case, one requires more measurements compared to our method to achieve the same probability of recovery for a given rank.

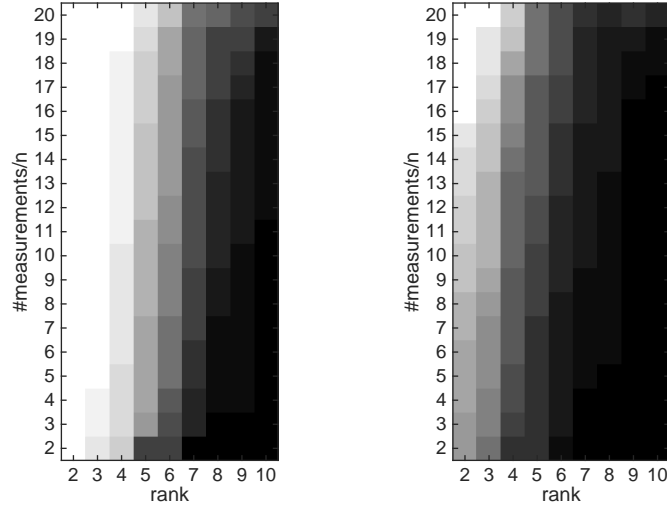
5.2 Tensor Completion: Phase Transition

We again considered tensors of size $30 \times 30 \times 30$, varied the rank of the tensors from 2 to 10, and obtained random measurements from four slices (without loss of generality we may assume they are the first 2 slices across modes 1 and 2). The number of measurements obtained varied as $n \times [2, 20]$. Figure 5.2 shows the phase transition plots of our method. We deem the method to be a “success” if the MSE of the recovered tensor is less than 10^{-5} . Results were averaged over 10 independent trials.

5.3 Speed Comparisons

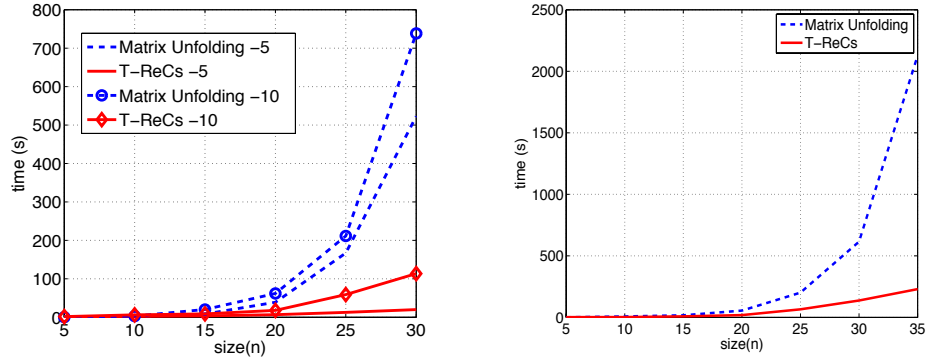
We finally compared the time taken to recover an $n \times n \times n$ tensor of rank 3. Figure 3(a) shows that, T-ReCs with four smaller nuclear norm minimizations is far more scalable computationally as compared to the method of unfolding the tensor to a large matrix and then solving a single nuclear norm minimization program. This follows since matricizing the tensor involves solving for an $n^2 \times n$ matrix. Our method can thus be used for tensors that are orders of magnitude larger than competing methods.

Along lines similar to the recovery case, we compared execution times to complete a $35 \times 35 \times 35$ sized tensor. Figure 3(b) shows again that the matrix completion approach takes orders of magnitude more time than that taken by our method. We average the results over 10 independent trials, and set $r = \frac{n}{5}$, $m = 3nr$



(a) Phase transition for tensor completion using T-ReCs. ($n=30$) (b) Phase transition for tensor completion using [35]. ($n=30$)

Figure 2: Phase transition plots for tensor recovery. Results are averaged over 10 independent trials. White indicates success whereas black indicates failure.



(a) Time taken to recover third order tensors. The numbers 5 and 10 in the legend refer to the cases where we obtain $5n$ and $10n$ measurements respectively

(b) Time taken for tensor completion by our method (T-ReCs) to that of flattening the tensor (Matrix Unfolding).

6 Conclusion and Future Directions

We introduced a computational framework for exact recovery of low rank tensors. A new class of measurements, known as *separable* measurements was defined, and sensing mechanisms of practical interest such as random projections and tensor completion with samples restricted to a few slices were shown to fit into the separable framework. Our algorithm, known as T-ReCs, built on the classical Leurgans' algorithm for tensor decomposition, was shown to be computationally efficient, and enjoy almost optimal sample complexity guarantees in both the random projection and the

completion settings. A number of interesting avenues for further research follow naturally as a consequence of this work:

1. **Robustness:** Our algorithm has been analyzed in the context of *noiseless* measurements. It would be interesting to study variations of the approach and the resulting performance guarantees in the case when measurements are noisy, in the spirit of the matrix completion literature [11].
2. **Non-separable measurements:** Our approach relies fundamentally on the measurements being separable. Tensor inverse problems, such as tensor completion in the setting when samples are obtained randomly and uniformly from the tensor do not fit into the separable framework. Algorithmic approaches for non-separable measurements thus remains an important avenue for further research.
3. **Tensors of intermediate rank:** Unlike matrices, the rank of a tensor can be larger than its (largest) dimension, and indeed increase polynomially in the dimension. The approach described in this paper addresses inverse problems where the rank is smaller than the dimension (low-rank setting). Extending these methods to the intermediate rank setting is an interesting and challenging direction for future work.
4. **Methods for tensor regularization:** Tensor inverse problems present an interesting dichotomy with regards to rank regularization. On the one hand, there is no known natural and tractable rank-regularizer (unlike the matrix case, the nuclear norm is not known to be tractable to compute). While various relaxations for the same have been proposed, the resulting approaches (while polynomial time), are neither scalable nor known to enjoy strong sample complexity guarantees. On the other hand, matrix nuclear norm has been used in the past in conjunction with matrix unfolding, but the resulting sample complexity performance is known to be weak. Our work establishes a third approach, we bypass the need for unfolding and expensive regularization, yet achieve almost optimal sample complexity guarantees and a computational approach that is also far more scalable. However, the method applies only for the case of separable measurements. This raises interesting questions regarding the need/relevance for tensor regularizers, and the possibility to bypass them altogether.

References

- [1] Anima Anandkumar, Rong Ge, Daniel Hsu, and Sham M Kakade. A Tensor Approach to Learning Mixed Membership Community Models. *arXiv.org*, February 2013.
- [2] Anima Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. preprint arXiv:1210.7559, 2012.
- [3] Anima Anandkumar, Rong Ge, and Majid Janzamin. Analyzing Tensor Power Method Dynamics: Applications to Learning Overcomplete Latent Variable Models. *arXiv.org*, November 2014.
- [4] A. Aswani. Positive Low-Rank Tensor Completion. *arXiv.org*, 2014.
- [5] B. Barak and A. Moitra. Tensor prediction, Rademacher complexity and random 3-XOR. preprint arXiv:1501.06521, 2015.

- [6] C F Beckmann and S M Smith. Tensorial Extensions of Independent Component Analysis for Multisubject fMRI Analysis. *NeuroImage*, 25(1):294–311, March 2005.
- [7] A. Bhaskara, M. Charikar, A. Moitra, and A. Vijayraghavan. Smoothed analysis of tensor decompositions. preprint arXiv:1311.3651, 2014.
- [8] Samuel Burer and Renato D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming (series B)*, 95:2003, 2001.
- [9] T. Tony Cai and Anru Zhang. ROP: Matrix recovery via rank-one projections. *Ann. Statist.*, 43(1):102–138, 02 2015.
- [10] E. Candes, M. Rudelson, T. Tao, and R. Vershynin. Error correction via linear programming. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 668–681, Oct 2005.
- [11] E.J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, June 2010.
- [12] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.
- [13] Emmanuel J Candès, J Romberg, and T Tao. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *Information Theory, IEEE Transactions on*, 52(2):489–509, February 2006.
- [14] Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *CoRR*, abs/1109.4499, 2011.
- [15] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.
- [16] Annie Chen. Context-Aware Collaborative Filtering System: Predicting the User’s Preference in the Ubiquitous Computing Environment. In *Location- and Context-Awareness*, pages 244–253. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [17] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. preprint arXiv:1306.2979, 2013.
- [18] Andrzej Cichocki. Era of Big Data Processing: A New Approach via Tensor Networks and Tensor Decompositions. *arXiv.org*, March 2014.
- [19] G. Dasarathy, P. Shah, B.N. Bhaskar, and R.D. Nowak. Sketching sparse matrices, covariances, and graphs via tensor products. *Information Theory, IEEE Transactions on*, 61(3):1373–1388, March 2015.
- [20] N. Goyal, S. Vempala, and Y. Xiao. Fourier PCA and robust tensor decomposition. preprint arXiv:1306.5825, 2014.

- [21] C. J. Hillar and L. H. Lim. Most tensor problems are NP-hard. *Journal of the ACM*, 60(6), 2013.
- [22] Prateek Jain and Inderjit S Dhillon. Provable inductive matrix completion. *arXiv preprint arXiv:1306.0626*, 2013.
- [23] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems 27*, pages 1431–1439. 2014.
- [24] Kaifeng Jiang, Defeng Sun, and Kim-Chuan Toh. Solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints. 69:133–162, 2013.
- [25] Jinman Kang, I Cohen, and G Medioni. Continuous multi-views tracking using tensor voting. In *Motion and Video Computing, 2002. Proceedings. Workshop on*, pages 181–186. IEEE Comput. Soc, 2002.
- [26] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. *Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering*. n-dimensional tensor factorization for context-aware collaborative filtering. ACM, New York, New York, USA, September 2010.
- [27] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *Information Theory, IEEE Transactions on*, 56(6):2980–2998, 2010.
- [28] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455, 2009.
- [29] J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Applicat.*, 18, 1977.
- [30] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *arXiv preprint arXiv:1410.6913*, 2014.
- [31] SE Leurgans, RT Ross, and RB Abel. A decomposition for three-way arrays. *SIAM Journal on Matrix Analysis and Applications*, 14(4):1064–1083, 1993.
- [32] Lek-Heng Lim and Pierre Comon. Multiarray Signal Processing: Tensor Decomposition Meets Compressed Sensing. *arXiv.org*, (6):311–320, February 2010.
- [33] Ji Liu, P Musialski, P Wonka, and Jieping Ye. Tensor Completion for Estimating Missing Values in Visual Data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):208–220, 2013.
- [34] Eduardo Martinez-Montes, Pedro A Valdés-Sosa, Fumikazu Miwakeichi, Robin I Goldman, and Mark S Cohen. Concurrent EEG/fMRI Analysis by Multiway Partial Least Squares. *NeuroImage*, 22(3):1023–1034, July 2004.
- [35] C. Mu, B. Huang, J. Wright, and D. Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. preprint arXiv:1307.5870, 2013.

- [36] E Papalexakis, U Kang, C Faloutsos, and N Sidiropoulos. Large Scale Tensor Decompositions: Algorithmic Developments and Applications. *IEEE Data Engineering Bulletin - Special Issue on Social Media*, 2013.
- [37] Nikhil Rao, Parikshit Shah, and Stephen Wright. Conditional gradient with enhancement and truncation for atomic-norm regularization. In *NIPS Workshop on Greedy Algorithms*, 2013.
- [38] Benjamin Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, December 2011.
- [39] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
- [40] Bernardino Romera-paredes, Hane Aung, Nadia Bianchi-berthouze, and Massimiliano Pontil. Multilinear multitask learning. In Sanjoy Dasgupta and David Mcallester, editors, *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, pages 1444–1452. JMLR Workshop and Conference Proceedings, May 2013.
- [41] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis. Parallel Factor Analysis in Sensor Array Processing. *Signal Processing, IEEE Transactions on*, 48(8):2377–2388, August 2000.
- [42] Age Smilde, Rasmus Bro, and Paul Geladi. *Multi-way Analysis: Applications in the Chemical Sciences*. John Wiley & Sons, 2005.
- [43] J Sun, S Papadimitriou, C Y Lin, N Cao, S Liu, and W Qian. MultiVis: Content-Based Social Network Exploration through Multi-way Visual Analysis. *SDM*, 2009.
- [44] Jimeng Sun, Dacheng Tao, and Christos Faloutsos. *Beyond Streams and Graphs: Dynamic Tensor Analysis*. dynamic tensor analysis. ACM, New York, New York, USA, August 2006.
- [45] Gongguo Tang, B N Bhaskar, P Shah, and B Recht. Compressed Sensing Off the Grid. *Information Theory, IEEE Transactions on*, 59(11):7465–7490, 2013.
- [46] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. preprint arXiv:1010.0789, 2011.
- [47] Hua Wang, Feiping Nie, and Heng Huang. Low-rank tensor completion with spatio-temporal consistency. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.
- [48] Svante Wold, Paul Geladi, Kim Esbensen, and Jerker Öhman. Multi-way principal components-and PLS-analysis. *Journal of Chemometrics*, 1(1):41–56, January 1987.
- [49] M. Yuan and C.-H. Zhang. On tensor completion via nuclear norm minimization. preprint arXiv:1405.1773, 2014.
- [50] Qiang Zhang, Han Wang, Robert J Plemmons, and V Paul Pauca. Tensor methods for hyperspectral data analysis: a space object material identification study. *Journal of the Optical Society of America A*, 25(12):3001–3012, December 2008.
- [51] Z. Zhang and S. Aeron. Exact tensor completion using t-SVD. preprint arXiv:1502.04689, 2015.